technical note technical note techn

User Request Evaluation Tool (URET) Conflict Probe Sensitivity to Weather Forecast Errors

Mike M. Paglione
Lori Charles
Shurong Liu
Robert D. Oaks
Hollis F. Ryan, Ph.D.
J. Scott Summerill

October 2002

DOT/FAA/CT-TN02/tbs

U.S. Department of Transportation
Federal Aviation Administration

William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

NOTICE

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

Technical Report Documentation Page

1. Report No. DOT/FAA/CT-TN02/tbs	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle User Request Evaluation Tool (URET) Conflict Probe Sensitivity to Weather Forecasts Errors		**5. Report Date** October 2002 6. Performing Organization Code ACB-330
7. Author(s) Mike M. Paglione, Lori Charles, Shurong Liu, Robert D. Oaks, Hollis F. Ryan, Ph.D., J. Scott Summerill		8. Performing Organization Report No. DOT/FAA/ CT-TN02/tbs
9. Performing Organization Name and Address U. S. Department of Transportation Federal Aviation Administration, William J. Hughes Technical Center Atlantic City International Airport, NJ 08405		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address U. S. Department of Transportation Free Flight Phase 1 Program Office, AOZ-200 Conflict Probe/URET Product Team Washington, D. C. 20590		13. Type of Report and Period Covered Technical Note 14. Sponsoring Agency Code AOZ-200

15. Supplementary Notes

The authors identified above represent the following organizations: Mike Paglione with FAA ACB-330; Lori Charles, Shurong Liu, Robert D. Oaks, Hollis F. Ryan, Ph.D., J. Scott Summerill with Veridian Corporation

16. Abstract

This study investigated the User Request Evaluation Tool's (URET) prediction sensitivity to weather forecast error. A quantitative experiment was designed and performed by the Federal Aviation Administration's Conflict Probe Assessment Team (CPAT) to evaluate the impact of weather forecast errors on URET trajectory and conflict predictions. The experiment used about two hours of traffic data recorded at the Indianapolis en route center in May 1999. The flights were time shifted to generate a sufficient number of test conflicts using a genetic algorithm technique developed by CPAT. This time-shifted scenario was used as input to the URET Prototype. To induce weather forecast error, the weather input file (Rapid Update Cycle, RUC) was altered by adding 20 or 60 knots to the wind magnitude, 45 or 90 degrees to the wind direction, and 5 or 15 degrees Kelvin to the air temperature. This produced seven URET runs for the experiment – the unaltered control run and six treatment runs. The analysis compared the control run against the treatment runs. A methodology was developed to compare the trajectory and conflict prediction accuracy of these runs. A statistical analysis provided evidence that the forecast errors in wind magnitude and direction had significant effect on the longitudinal trajectory error and a modest impact on retracted false alerts, which caused at most an increase of 0.06 in the false alert probability. It also showed that the air temperature runs did not have a significant effect.

Based on this experiment, a controller suspecting errors in the input wind forecast should expect only a modest impact on URET predictions. The impact would mainly be a moderate increase in the number of retractions of its conflict predictions (defined in this study as a retracted false alerts). If the controller notices an increase in retractions, it may be symptomatic of inaccurate wind forecasts, which should be investigated.

17. Key Words URET, User Request Evaluation Tool Free Flight Conflict Prediction Accuracy Missed, False, and Valid Alerts Horizontal Trajectory Error Longitudinal Trajectory Error Weather Forecast Error RUC, Rapid Update Cycle Trajectory Accuracy DST, Decision Support Tool HCS, Host Computer System Lateral Trajectory Error Vertical Error Trajectory Error Genetic Algorithm	**18. Distribution Statement** This report is approved for public release and is on file at the William J. Hughes Technical Center, Aviation Security Research and Development Library, Atlantic City International Airport, New Jersey 08405. This document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161.

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	122	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

Acknowledgements

We thank Bill Dudzik and Gary Wright at AST Engineering Services for providing the initial guidance on conversion of the weather forecast files and sharing their software tools with us.

MITRE Center for Advanced Aviation System Development (CAASD) personnel, in particular Bill Arthur and John Mayo, were also very helpful in providing instructions on the use of the URET Prototype's recording and parsing tools. Also, they provided several of the MITRE references used in the literature survey of this study.

Furthermore, a special thanks goes to our sponsors in the FAA's Free Flight Phase One Office (AOZ-200), Tom Spellerberg, and in AUA-200, Jesse Wijntjes. They supplied the encouragement and funding for this research to take place.

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

Executive Summary

This study investigated the User Request Evaluation Tool's (URET) prediction sensitivity to weather forecast error. A quantitative experiment was designed and performed by the Federal Aviation Administration's Conflict Probe Assessment Team (CPAT) to evaluate the impact of weather forecast errors on URET trajectory and conflict predictions. A literature survey on previous research in the weather forecast errors was conducted. Much of the research reported on methods to reduce weather forecast errors. MITRE research concluded that URET provides valuable conflict alert information in the absence of wind data. The major effect reported was a modest increase in the number of marginal conflict alerts. However, there was limited research on the sensitivity of URET's conflict prediction accuracy to weather forecast errors. This CPAT study provides a comprehensive analysis on their impact on both URET's trajectory and conflict prediction accuracy.

The experiment used about two hours of traffic data recorded at the Indianapolis en route center in May 1999. The flights were time shifted to generate a sufficient number of test conflicts using a genetic algorithm technique developed by CPAT. This time-shifted scenario was used as input to the URET Prototype. To induce weather forecast error, the weather input file (Rapid Update Cycle, RUC) was altered by adding 20 or 60 knots to the wind magnitude, 45 or 90 degrees to the wind direction, and 5 or 15 degrees Kelvin to the air temperature. This produced seven URET runs for the experiment – the unaltered control run and six treatment runs.

The analysis compared the control run against the treatment runs. A methodology was developed to compare the trajectory and conflict prediction accuracy of these runs. A statistical analysis provided evidence that the forecast errors in wind magnitude and direction had significant effect on the longitudinal trajectory error and a modest impact on retracted false alerts, which caused at most an increase of 0.06 in the false alert probability. It also showed that the air temperature runs did not have a significant effect.

Four flights and their encounters with other aircraft were analyzed in detail to help determine the causes of this overall effect. These analyses revealed that the error added to the forecasted wind data causes additional errors in predicted positions. The new errors are principally along the flight path or longitudinal errors caused by inaccurate predictions of ground speed. The increase in longitudinal error is consistent with the trajectory accuracy results and the statistical analysis. Vertical position errors are caused primarily by errors in predicted climb rate. The predicted climb rates are affected only slightly by errors in predicted winds, while the predicted climb angles are affected somewhat more. This small vertical effect was not statistically significant. The position errors cause URET to rebuild trajectories more frequently resulting in retractions in conflict predictions. Thus, the number of retracted false alerts is increased. The last flight example demonstrates that an individual flight may be greatly impacted, but the aggregate effect on missed alert probability was not statistically significant. The URET trajectory reconformance logic correctly adjusted its trajectories to avoid missing conflict predictions. As expected, this same reconformance logic caused more retracted false alerts to be generated.

Operationally, weather forecasts may be inaccurate due to the presence of highly dynamic weather or outages in the interfaces to the National Weather Service. Based on this experiment, a controller suspecting errors in the input wind forecast should expect only a modest impact on URET predictions. The impact would mainly be a moderate increase in the number of retractions of its conflict predictions (defined in this study as a retracted false alerts). If the controller notices

an increase in retractions, it may be symptomatic of inaccurate wind forecasts, which should be investigated.

Although the effect of weather errors on URET Prototype's predictions was shown to be minor, future research should confirm the applicability of these results to the production version of URET (Core Capability Limited Deployment, CCLD). In addition, future research should investigate the impact of convective weather on URET predictions.

Table of Contents

EXECUTIVE SUMMARY ..... **IX**

TABLE OF CONTENTS ..... **XI**

LIST OF FIGURES ..... **XIV**

LIST OF TABLES ..... **XV**

1 INTRODUCTION **1**

1.1 PURPOSE 1
1.2 BACKGROUND 1
1.3 SCOPE 1
1.4 DOCUMENT ORGANIZATION 2

2 STUDY METHODOLOGY **3**

2.1 DESCRIPTION OF EXPERIMENT 3
2.1.1 Review of Related Studies 3
2.1.2 Design of Experiment 5
2.1.3 Input Data to Experiment 6
2.1.3.1 Input Air Traffic Scenario 6
2.1.3.1.1 Input Air Traffic Messages 6
2.1.3.1.2 Overview of the Scenario Generation Process 7
2.1.3.1.2.1 Data Extraction 7
2.1.3.1.2.2 Data Modification 8
2.1.3.1.2.3 Scenario Generation 9
2.1.3.1.3 Summary of Genetic Algorithm Run 9
2.1.3.1.3.1 Input Parameters 9
2.1.3.1.3.2 Analysis of Run 10
2.1.3.2 Input Weather Files 13
2.1.3.2.1 Description of RUC files 13
2.1.3.2.2 Description of Software Tools Used to Process Weather Data 19
2.1.3.2.2.1 Software Developed by the National Weather Service 19
2.1.3.2.2.1.1 wgrib 19
2.1.3.2.2.1.2 gribw 19
2.1.3.2.2.2 Software Developed by CPAT 19
2.1.3.2.2.2.1 getRUC 19
2.1.3.2.2.2.2 modRUC 20
2.1.4 Output Data of Experiment 21
2.2 DESCRIPTION OF ANALYSIS 22
2.2.1 Scenario Processing 22
2.2.2 Trajectory Accuracy Measurement 22
2.2.2.1 Trajectory Error Deviations Between Runs 23
2.2.2.2 Trajectory Accuracy Statistical Methods 23
2.2.2.2.1 Preliminary Data Analysis 23
2.2.2.2.2 Design of Experiment Methodology 24
2.2.2.2.3 Nonparametric Statistical Experiment 25
2.2.2.2.4 SAS Univariate Statistical Procedure 26

2.2.2.2.5 Graphical Statistical Plots 26
2.2.2.2.6 Applied Statistical Analysis 28
2.2.3 Conflict Prediction Accuracy Measurement 31
2.2.3.1 Missed, Valid, and False Alerts 32
2.2.3.1.1 Methodology of Measuring Missed and False Alerts 33
2.2.3.1.2 Reason Codes for Missed, False, and Valid Alerts 35
2.2.3.2 Conflict Prediction Run Comparison 36
2.2.3.2.1 Statistical Significance of Missed and False Alert Differences 41
2.2.3.2.2 Combinations of Conflict Prediction Run Comparisons 43
2.2.3.3 Valid Alert Attribute Comparison 43
2.2.3.3.1 Definition of Warning and Start Time Comparisons 44
2.2.3.3.2 Statistical Tests for Warning and Start Time Comparisons 45

3 ANALYSIS AND RESULTS 47

3.1 TRAJECTORY ACCURACY RESULTS 47
3.1.1 Wind Magnitude 47
3.1.1.1 Statistical Results 47
3.1.1.2 General Observations 49
3.1.2 Wind Direction 49
3.1.2.1 Statistical Results 49
3.1.2.2 General Observations 50
3.1.3 Air Temperature 51
3.1.3.1 Statistical Results 51
3.1.3.2 General Observations 52
3.1.4 Trajectory Stability Results 53
3.1.5 Summary of Trajectory Accuracy Results 54
3.2 CONFLICT PREDICTION RESULTS 55
3.2.1 Individual Run's Missed, Valid, and False Alert Counts 56
3.2.1.1 Analysis AA's Individual Run Conflict Prediction Results 56
3.2.1.2 Analysis BA's Individual Run Conflict Prediction Results 57
3.2.1.3 Analysis AB's Individual Run Conflict Prediction Results 58
3.2.1.4 Analysis BB's Individual Run Conflict Prediction Results 59
3.2.2 Results on Conflict Prediction Run Comparison 60
3.2.2.1 Analysis AA's Run Comparison Conflict Prediction Results 60
3.2.2.2 Analysis BA's Run Comparison Conflict Prediction Results 63
3.2.2.3 Analysis AB's Run Comparison Conflict Prediction Results 65
3.2.2.4 Analysis BB's Run Comparison Conflict Prediction Results 68
3.2.3 Results on Comparison of Common Valid Alert Attributes 70
3.2.4 Summary of Conflict Prediction Accuracy Results 72
3.3 FLIGHT EXAMPLES 73
3.3.1 Flight Example #1 73
3.3.1.1 Flight Path 73
3.3.1.2 Forecasted Wind Runs 73
3.3.1.3 Trajectories 77
3.3.1.4 Ground Speed 77
3.3.1.5 Climb Angle 78
3.3.1.6 Climb Rate 78
3.3.1.7 Forecasted Wind Direction Runs 80
3.3.1.8 Forecasted Air Temperature Runs 80
3.3.1.9 Conflict Predictions 80
3.3.1.10 Summary of Flight Example #1 81

3.3.2 Flight Example #2 81
3.3.2.1 Flight Path 81
3.3.2.2 Description of the Encounter 81
3.3.2.3 Forecasted Wind Runs 82
3.3.2.4 Forecasted Air Temperature Runs 82
3.3.2.5 Longitudinal Errors 82
3.3.2.6 Conflict Predictions 82
3.3.2.7 Summary of Flight Example #2 83
3.3.3 Flight Example #3 85
3.3.3.1 Flight Path 85
3.3.3.2 Description of the Encounter 85
3.3.3.3 Forecasted Wind Runs 85
3.3.3.4 Forecasted Air Temperature Runs 85
3.3.3.5 Longitudinal Errors 86
3.3.3.6 Conflict Predictions 86
3.3.3.7 Summary of Flight Example #3 86
3.3.4 Flight Example #4 89
3.3.4.1 Flight Paths 89
3.3.4.2 Conflict 90
3.3.4.3 Forecasted Wind Runs 90
3.3.4.4 Forecasted Air Temperature Runs 90
3.3.4.5 Longitudinal Errors 90
3.3.4.6 Conflict Predictions 90
3.3.4.6.1 Control Run 90
3.3.4.6.2 Forecasted Wind Runs 93
3.3.4.6.3 Forecasted Air Temperature Runs 93
3.3.4.7 Summary of Flight Example #4 93
3.3.5 Summary of Flight Example Results 94

4 CONCLUSION 95

4.1 OVERVIEW OF THE EXPERIMENT 95
4.2 STATISTICAL ANALYSIS CONCLUSIONS 95
4.3 FLIGHT ANALYSIS CONCLUSIONS 96
4.4 OPERATIONAL RECOMMENDATIONS 96
4.5 FUTURE RESEARCH 96

5 LIST OF ACRONYMS 97

6 REFERENCES 99

7 INDEX 103

APPENDICES 107

List of Figures

Figure 2.1-1: Scenario Generation Process 7
Figure 2.1-2: Fitness vs. Generation Number 11
Figure 2.1-3: Standard Deviation Fitness vs. Generation Number 11
Figure 2.1-4: Number of Conflicts vs. Generation Number 12
Figure 2.1-5: Distribution of Time Shifts 13
Figure 2.1-6: Modification of Wind Vector 21
Figure 2.2-1: Q-Q Plots of Horizontal and Vertical Trajectory Deviation Error 24
Figure 2.2-2: Components of JMP Quantile Box Plot 27
Figure 2.2-3: Example JMP Box Plot of Horizontal Errors for Comparison 000-200 27
Figure 2.2-4: Plot of Median Difference in Horizontal Error by Look-Ahead Time for Wind Magnitude during Level Flight 28
Figure 2.2-5: Process A - Valid and Missed Alert Processing 34
Figure 2.2-6: Process B - False Alert Processing 35
Figure 3.2-1: Analysis AA Run Percentage Difference of Standard and Retracted False Counts 62
Figure 3.2-2: Analysis BA Run Percentage Difference of Standard and Retracted False Counts. 65
Figure 3.2-3: Analysis AB Run Percentage Difference of Standard and Retracted False Counts. 67
Figure 3.2-4: Analysis BB Run Percentage Difference of Standard and Retracted False Counts. 70
Figure 3.3-1 Example 1 Route 74
Figure 3.3-2 Example 1 Vertical Profiles 75
Figure 3.3-3 Example 1 Relative Wind Along Track 76
Figure 3.3-4 Example 1 Track and Wind Directions Along Track 76
Figure 3.3-5 Example 1 Cumulative Along Track and Trajectory Distances – Control Run 77
Figure 3.3-6 Example 1 Cumulative Along Track and Trajectory Distances – Run 200 78
Figure 3.3-7 Example 1 Angle of Climb - Trajectories Built at 69043 79
Figure 3.3-8 Example 1 Rate of Climb - Trajectories Built at 69043 79
Figure 3.3-9 Example 2 RMS Values of the Differences in Relative Wind Magnitudes Between Runs 83
Figure 3.3-10 Example 2 Standard Deviations of the Differences in Longitudinal Errors between Runs 84
Figure 3.3-11 Example 3 RMS Values of the Differences in Relative Wind Magnitudes Between Runs 87
Figure 3.3-12 Example 3 Standard Deviations of the Differences in Longitudinal Errors Between Runs 87
Figure 3.3-13 Example 4 RMS Values of the Differences in Relative Wind Magnitudes Between Runs 91
Figure 3.3-14 Example 4 Standard Deviations of the Differences in Longitudinal Errors between Runs 91

List of Tables

Table 2.1-1: Experiment Control and Treatment Combinations 6
Table 2.1-2: Air Traffic Database 8
Table 2.1-3: Input Constraint Bin Bounds 10
Table 2.1-4: Constraint Bin Conflict Counts 12
Table 2.1-5: RUC 211 Grid Description 14
Table 2.1-6: Availability of RUC 211 Files and Number of Forecast Hours 15
Table 2.1-7: Geopotential Height Data in the RUC 211 Files (HGT) 16
Table 2.1-8: Temperature Data in the RUC 211 Files (TMP) 17
Table 2.1-9: North-South Component of the Wind Data in the RUC 211 Files (VGRD) 18
Table 2.1-10: Description of the Weather Table 20
Table 2.2-1: Count and Percentage of Observations Retained by Phase-of-Flight and Look-Ahead Time 29
Table 2.2-2: Quantiles for Run Set 000-100, Lateral Error and Level Flight 29
Table 2.2-3: Test Results for Run Set 000-100, Lateral Error and Level Flight 30
Table 2.2-4: Quantiles for Vertical Error Run Set 000-100 and In-Transition Flight 30
Table 2.2-5: Test Results for Run Set 000-100, Vertical Error and In-Transition Flight 31
Table 2.2-6: URET Alert and Conflict Event Combinations 33
Table 2.2-7: Reason Codes for Conflict Prediction Types 36
Table 2.2-8: Comparison of Two Run's Resulting Alert and Conflict Event Combinations 37
Table 2.2-9: Comparison of Two Run's Resulting Alert and Conflict Event Combinations With Discard Events 38
Table 2.2-10: Summary of Event Count Variables 39
Table 2.2-11: Additional Comparison Probabilities 41
Table 2.2-12: Conflict Prediction Comparison Program Evaluation Codes 43
Table 3.1-1: Wind Magnitude Comparison Results for Level Flight 48
Table 3.1-2: Wind Magnitude Comparison Results for In-Transition Flight 48
Table 3.1-3: Wind Direction Comparison Results for Level Flight 50
Table 3.1-4: Wind Direction Comparison Results for In-Transition Flight 50
Table 3.1-5: Air Temperature Comparison Results for Level Flight 52
Table 3.1-6: Air Temperature Comparison Results for In-Transition Flight 52
Table 3.1-7: Total Trajectory Count Per Run 54
Table 3.1-8: Trajectory Count Run Comparison Statistics 54
Table 3.2-1: Conflict Prediction Counts and Error Probabilities for Analysis AA 56
Table 3.2-2: Conflict Prediction Reason Code Counts for Analysis AA 57
Table 3.2-3: Conflict Prediction Counts and Error Probabilities for Analysis BA 57
Table 3.2-4: Conflict Prediction Reason Code Counts for Analysis BA 58
Table 3.2-5: Conflict Prediction Counts and Error Probabilities for Analysis AB 58
Table 3.2-6: Conflict Prediction Reason Code Counts for Analysis AB 59
Table 3.2-7: Conflict Prediction Counts and Error Probabilities for Analysis BB 59
Table 3.2-8: Conflict Prediction Reason Code Counts for Analysis BB 60
Table 3.2-9: Conflict Prediction Run Comparison Statistics for Analysis AA 61
Table 3.2-10: Analysis AA False Alert Event Statistical Test 62
Table 3.2-11: Analysis AA Missed Alert Event Statistical Test 62
Table 3.2-12: Conflict Prediction Run Comparison Statistics for Analysis BA 63
Table 3.2-13: Analysis BA False Alert Event Statistical Test 64
Table 3.2-14: Analysis BA Missed Alert Event Statistical Test 64

Table 3.2-15: Conflict Prediction Run Comparison Statistics for Analysis AB 66
Table 3.2-16: Analysis AB False Alert Event Statistical Test 67
Table 3.2-17: Analysis AB Missed Alert Event Statistical Test 67
Table 3.2-18: Conflict Prediction Run Comparison Statistics for Analysis BB 68
Table 3.2-19: Analysis BB False Alert Event Statistical Test 69
Table 3.2-20: Analysis BB Missed Alert Event Statistical Test 69
Table 3.2-21: Comparison of Common Valid Alert's Warning Time... 71
Table 3.2-22: Comparison of Common Valid Alert's Predicted Conflict Start Time 71
Table 3.3-1 Example 1 Notification Sets for ABC1000 with Other Flights 80
Table 3.3-2 Example 2 Notification Sets for Encounter with Flights DEF1000 and TUV1000 ... 84
Table 3.3-3 Example 3 Notification Sets for Encounter with Flights GHI1000 and QRS1000 – Adherence Rule Applied 88
Table 3.3-4 Example 3 Notification Sets for Encounter with Flights GHI1000 and QRS1000 - Adherence Rule Ignored 89
Table 3.3-5 Example 4 Notification Sets for Conflict with Flights JKL1000 and NOP1000........ 92

1 Introduction

1.1 Purpose

This technical note documents the results of an independent analysis on the User Request Evaluation Tool's (URET) sensitivity to weather forecast error. The ACB-330 Conflict Probe Assessment Team (CPAT) conducted this study at the Federal Aviation Administration (FAA) William J. Hughes Technical Center (WJHTC)[1]. In this analysis an experiment was designed and performed with induced weather forecast errors. Quantitative measures of trajectory and conflict prediction accuracy were applied to determine the impact of the induced weather forecast errors. Statistical hypothesis tests, aggregate point statistics, and graphical techniques were employed to investigate the effects of weather forecast errors on URET predictions. Finally, selected flight examples are presented, which demonstrate URET sensitivity to weather forecast errors.

1.2 Background

To achieve the FAA's goals of Free Flight, advances in ground and airborne automation are required. One of the most important ground based tools is a conflict detection tool or conflict probe (CP). A CP is a decision support tool (DST) that provides the air traffic controller with predictions of conflicts, or loss of minimum separation between a pair of aircraft or between an aircraft and protected airspace, for a parametric time into the future, typically 20 minutes. A CP predicts the flight path of an aircraft, continuously monitors that flight path from current aircraft position information, and probes for conflicts with other aircraft and incursions into restricted airspace. A CP makes these predictions based on air traffic control clearances, radar surveillance position reports, aircraft and airspace characteristic data, and weather forecasts. Therefore, inaccuracies in this input data is expected to cause error in the predictions the conflict probe makes.

The FAA Free Flight Office (AOZ-200) tasked CPAT to examine the sensitivity of the FAA's en route CP, known as URET, on one of its input sources, specifically the weather forecasts. To accomplish this, CPAT developed tools to induce weather forecast errors on the input weather data. The objective of the analysis was to determine URET's trajectory and conflict prediction sensitivity to degraded weather forecasts.

1.3 Scope

This technical note is provided as a reference report and documents the experiment, methodology used, analysis results, and presents examples of specific flights. Using the MITRE Center for Advanced Aviation System Development (CAASD) Prototype version of URET, the experiment altered three factors in weather forecast errors: wind magnitude, wind direction, and air temperature. A literature review was performed that helped determine the level of these factors. The experiment was composed of one URET run, referred to as the control run, with unaltered forecast weather data, and then was repeated with the six altered weather forecast files, referred to as the treatment runs. For the treatment runs, each of these three factors was altered twice with one low and one high level. The analysis compares the URET trajectory and conflict prediction accuracy of the treatment runs to the control run. The trajectory and conflict prediction errors

[1] In 1996 the Federal Aviation Administration's Traffic Flow Management Branch (ACT-250) established the Conflict Probe Assessment Team (CPAT) to evaluate the accuracy of conflict probes. In 2002, CPAT became a part of the Simulation and Modeling Group (ACB-330).

induced in the experiment for the selected factor levels were evaluated. The effects of each of the three factors were considered separately. Interactions between the factors were not investigated.

1.4 Document organization

This technical note is organized into four primary sections and four appendices. Section 2 provides a detailed description of the experiment and analysis methodology used in the study. Section 3 provides the trajectory and conflict prediction statistical results of the experiment and presents several flight examples, illustrating the causes of the prediction errors. Section 4 provides the conclusion of the experimental results. This technical note also includes document references, a list of acronyms, and a comprehensive subject index. In addition, four appendices provide detailed statistical charts and graphs. Appendix A and B presents box and median plots of the trajectory prediction results, respectively. Appendix C presents box plots of each weather forecast factor considered in the experiment. Appendix D presents the detailed analysis of the first flight example, which is presented partially in Section 3.3.1.

2 Study Methodology

2.1 Description of Experiment

This section provides a detailed description of the data, processes, and tools used to perform the study's quantitative experiment using a prototype version of URET, referred to as the URET Prototype built by MITRE CAASD. First, a review of previous studies is presented and then the design of the experiment is discussed. Next, the details of URET's generated input data and resulting output data are presented.

2.1.1 Review of Related Studies

Weather forecast error and its effect on URET is not a novel research topic. Several organizations and researchers have investigated the impact of weather forecast errors. However, no one has performed an experiment and applied a comprehensive analysis to determine both URET's trajectory and conflict prediction's sensitivity to weather forecast error. The following paragraphs highlight the findings the other researchers have reported.

MITRE CAASD, the developers of URET, performed the most comparable research to this CPAT study in [Lindsay, 1997b]. In the [Lindsay, 1997b], the sensitivity of the URET Prototype to weather data was measured by running the Algorithmic Evaluation Capability (AEC) version of URET (i.e. a simulation version of URET) using a five-hour air traffic scenario with and without its input weather forecast files. The conflict alerts generated and their predicted warning times (the time intervals between the posting of the alerts and their predicted conflict start times) for the two runs were compared. The trajectory accuracy and reconformance[2] rates were also compared. It was found that the lack of weather data increased the longitudinal track-to-trajectory deviations at large look-ahead times and the lateral and vertical deviations were relatively unchanged. The predicted warning times for the alerts common to both runs increased slightly. Alerts were generated by the no-wind run, which were not generated by the baseline run and vice versa. The trajectory reconformance rate went up slightly. It was concluded that URET can provide valuable conflict alert information in the absence of weather forecast data and the major effect was a modest increase in the number of marginal conflict alerts. The study did not examine these conflict predictions in terms of their accuracy degradation in the absence of the weather forecasts but reported that the quantity of predictions increased as a result and inferred they were caused by increases in the longitudinal track-to-trajectory deviations.

In [Cole et al., 2000], a collaborative effort of researchers from Massachusetts Institute of Technology Lincoln Laboratory (MIT/LL), National Aeronautics Space Administration (NASA) Ames Research Center, and National Oceanographic and Atmospheric Administration (NOAA) Forecast Systems Laboratory (FSL) reported on a year-long weather study. The data was collected over the Denver Air Route Traffic Control Center (ARTCC) airspace. The study was conducted to better understand wind prediction errors, to establish metrics for quantifying large wind prediction errors, and to validate two approaches to improve wind prediction accuracy. Besides an exhaustive analysis of 13 months of wind prediction data from the Rapid Update Cycle (RUC) forecasts and the Aircraft Communication Addressing and Reporting System

[2] Trajectory reconformance is defined as URET's method of monitoring and rebuilding its aircraft trajectory predictions when the current reported track position is outside the trajectory's conformance bounds. The conformance bounds are regions of uncertainty built around the trajectory centerline. The more URET rebuilds or reconforms a trajectory indicates its uncertainty in its trajectory prediction.

(ACARS), a series of aircraft flight tests were also performed. The on-average wind prediction accuracy was reported to be sufficient, but the analysis revealed that occasionally large errors existed over large regions of airspace. It was concluded that these large errors were present sufficiently to degrade the operational acceptance of DST predictions. One key result of the flight tests reported that the wind prediction error caused the greatest impact to trajectory prediction error at a look-ahead time of 20 minutes. Furthermore, two approaches were presented that improved the original RUC wind predictions and greatly reduced the occurrence of these large wind prediction errors. Therefore, the research in [Cole et al., 2000] provided insight into the wind prediction errors and guidance on realistic error levels to investigate for the CPAT's study on URET. It also supplied further evidence on the impact wind error has on DST predictions.

In another study, documented in [Wanke, 1997], MITRE CAASD evaluated the use of aircraft speed and wind reports to reduce trajectory prediction errors in URET algorithms. The reports were obtained from the aircraft in flight via ACARS and added to the trajectory modeling process. It was found that the aircraft reports improved the trajectory longitudinal prediction error by an average of 10% to 15%. The number of trajectory reconformances was also reduced. The ACARS reported data was used to create a statistical model of the airspeed and wind variations. Therefore, the research in [Wanke, 1997] provided insight into the wind errors themselves as well as the impact on URET's trajectory predictions. However, it had provided no analysis on the impact on URET's conflict prediction accuracy, which is a major emphasis in this CPAT study.

In [Schwartz and Benjamin, 1998], the researchers from NOAA compared the accuracy of the RUC-1 and the newer RUC-2 weather forecasts. RUC-2 has higher resolution, a one hour assimilation cycle rather than a three hour assimilation cycle, more input data, and better physical models. The actual winds aloft were obtained from aircraft in flight via the ACARS data link. The differences between the observed winds aloft (from ACARS) and the predicted winds aloft (from RUC) were used to calculate along path distance prediction errors and errors in the predicted times of arrival. An analysis of 140,000 flights, collected over a 13-month period, found that 15 minute en route segments accumulated time of arrival errors of 15 seconds and that 15 minute ascent/descent route segments accumulated time of arrival errors of eight seconds. The focus of this report is on the quality of the weather data, which once again provides insight into the underlying accuracy of these weather forecasts. This research does provide some analysis on the effects of the weather forecast errors on trajectory predictions and none on the sensitivity to a DST's aircraft conflict predictions.

In yet another study, [Sherry, 1999], MITRE CAASD reported on the accuracy lost in forecasted winds aloft when the data is provided in a resolution below what is available with either RUC-1 or RUC-2. A MITRE tool known as Winds Aloft Require Evaluation System (WARES) was presented and used to filter bad data from RUC, ACARS, and Meteorological Data Collection and Reporting System (MDCRS) data sources. The experiment paired aircraft wind reports (along and cross-track wind vector components) with the forecasted reports, filtered erroneous observations and statistically compared the difference. The study presented the specifics of twenty independent experiments that corresponded to the combinations of data resolution and forecast intervals available with RUC-1 and RUC-2. The statistical analysis used the Root Mean Square Wind Vector Error (RMSWVE) which is the standard RMS statistic employed by WARES. The study concluded coarser wind models like RUC-1 relative to RUC-2 can reduce the random noise in the wind aloft forecasts and consequently offset any loss of accuracy due to the decreased resolution. Thus, the study deduced that resolution based requirements for gridded-forecast weather data (like RUC) do not necessarily provide the best available accuracy regarding winds aloft prediction. This research documented in [Sherry, 1999] provides a thorough background into URET's input weather forecast files (RUC files) and thus this CPAT study. However, it does not provide any analysis of the impact for URET's predictions and suggests this as a future research area.

In conclusion, these references provide an extensive foundation from which the FAA CPAT weather sensitivity study is applied. The references present detailed descriptions and performance data on the existing weather forecast products and in some cases offer improved solutions for the future. Since the weather products are their primary focus, they only indirectly examined the impact on DSTs like URET. The study documented in [Lindsay, 1997b] was the exception. It directly examined URET sensitivity to the absence of timely weather forecasts, but the impact focused mainly on URET trajectory accuracy and only partially on URET conflict prediction accuracy. Therefore, this section's review of related literature provided further justification of performing a comprehensive analysis on the impact on both URET's trajectory and conflict prediction accuracy. In the more recent MITRE CAASD study documented in [Sherry, 1999], it was concluded: *"Future research will include performing sensitivity analysis of Air Traffic Management (ATM) automation to winds aloft error."* This is precisely the objective of the CPAT weather study documented here and presented first in the next section.

2.1.2 Design of Experiment

The focus of this study is to investigate URET's prediction sensitivity to weather forecast error. To examine these errors, a quantitative experiment was developed. The objective of the experiment was to evaluate what impact weather forecast errors have on URET trajectory and conflict predictions, if any, and determine whether or not the impact is statistically significant. To understand this phenomenon, wind and air temperature forecast errors were induced by altering URET's input weather forecast files.

The experiment consisted of extracting traffic data from Indianapolis ARTCC field recordings made on May 26, 1999. Two hours of traffic data was extracted and time-shifted to generate a scenario with a total of 211 aircraft-to-aircraft conflicts (see Section 2.1.3.1 for details). The experiment used the same traffic scenario throughout and only altered the weather forecast files.

The process of altering the weather forecast files and the tools to perform them will be presented in detail in Section 2.1.3.2. Briefly, the forecasted weather data was obtained from the National Weather Service (NWS) for same day in May 1999. This forecast data was formatted as Rapid Update Cycle 2 (RUC-2) gridded-binary files. As the main input source for the experiment, these files were modified throughout in wind magnitude, wind direction, and air temperature. The control run had no RUC file modifications, while all treatment runs had modified RUC files. The URET Prototype was run with the same air traffic scenario and these modified weather files in single center operation.

The three weather factors were altered individually at two different levels. The selection of these factors and levels were chosen based on research presented in [Wanke, 1997] and [Cole et al., 2000] and an internal empirical study on the control run RUC file. For example, in [Cole et al., 2000] a wind magnitude error of up to 60 knots was observed in a year-long study over Denver Air Route Traffic Control Center (ARTCC) airspace. In CPAT's analysis of the RUC file from May of 1999, a software tool was developed and implemented that extracted the wind magnitude, wind direction, and air temperature for each flight's Host Computer System (HCS) reported positions. The resulting weather forecasts were then extracted and summarized in box plots presented in Appendix C. The level one or low level was selected to cover approximately 50 percent of the data range. The level two is a higher value selected to cover most of the data range. Hence, wind magnitude was modified by adding 20 knots or 60 knots to all the forecasted winds. Similarly, wind direction was modified by adding 45 degrees or 90 degrees. Air temperature was modified by adding 5 degree Kelvin or 15 degree Kelvin to all the temperature forecast grid-points. This resulted in a total of seven URET runs; the one control run and six treatment runs. Table 2.1-1 lists these seven runs and their assigned run codes. These run codes are used throughout this document to refer to the associated URET run. The analysis compares each treatment run against the control run and in some cases the other treatment run in its category. For example, for the wind magnitude factor the control Run 000 is compared to the wind magnitude run with 20 knots added, Run 100, and the Run 200 with 60 knots added. For this example, the comparisons would be listed as 000-100 and 000-200. In some cases, the 100-200 will also be explored.

Table 2.1-1: Experiment Control and Treatment Combinations

Factor	Level	Run Code
Control Run	No change to RUC file	000
Wind Magnitude	Add 20 knots	100
	Add 60 knots	200
Wind Direction	Add 45 degrees	010
	Add 90 degrees	020
Air Temperature	Add 5 degrees Kelvin	001
	Add 15 degrees Kelvin	002

2.1.3 Input Data to Experiment

This section describes the input data for the experiment. There are two general sources of input: an input air traffic scenario and a weather forecast file. The input traffic scenario contains the recorded HCS messages (e.g. flight plan and surveillance position reports). The weather forecast files are used to build the proper aircraft trajectory profiles, taking into account the wind and air temperature dynamics of the atmosphere in which these aircraft fly. Both are essential to the proper operation of URET, so the details involved with supplying these input files to URET are a critical part of the experiment. The following subsections describe these input data sources and tools developed to generate them.

2.1.3.1 Input Air Traffic Scenario

2.1.3.1.1 Input Air Traffic Messages

An air traffic scenario is a data file describing the flow of aircraft traffic over a period of time. Scenario files contain time-stamped planning and advisory information and track data for the aircraft. The planning and advisory information describe the aircraft's planned flight; which includes its flight plan and flight plan amendments, interim altitude clearances, and hold information. The track data represents the aircraft's actual flight path. It consists of several fields including the flight's time-stamped horizontal coordinates and altitude. The scenario file used in this study is an ASCII file in a format compatible with a number of MITRE tools. (See [Lindsay, 1997b] and [Lindsay, 1998].)

2.1.3.1.2 Overview of the Scenario Generation Process

The scenarios used for this study are based on data recorded at the Indianapolis ARTCC (ZID) on May 26, 1999. They were created using software developed by CPAT. The individual flights within these scenarios follow actual flight routes, yet the air traffic in the scenarios contain aircraft-to-aircraft conflicts that do not exist in the field. An overview of this software is presented in [Oaks and Paglione, 2001], some of which is also presented in this document.

As depicted in Figure 2.1-1, the scenario generation process consists of three basic steps: data extraction, data modification, and scenario generation.

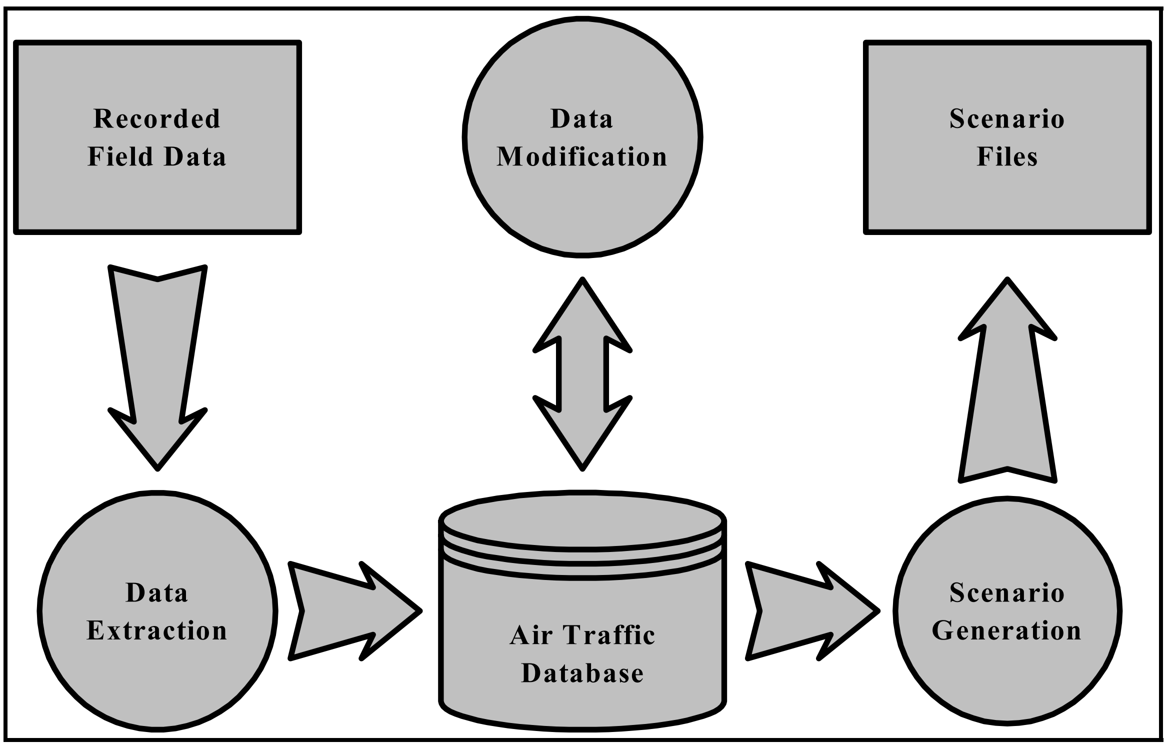


Figure 2.1-1: Scenario Generation Process

2.1.3.1.2.1 Data Extraction

The first step of the scenario generation process uses software to extract the data from the field recordings and populate an Air Traffic Database on an Oracle V8.1.6 database management system. This database consists of the twelve tables presented in Table 2.1-2. These tables are grouped into three categories: environmental tables, bookkeeping tables, and flight-centric tables. The environmental tables contain center specific and operational data. This includes coordinate conversion constants and preferential routing, sector assignment, and route status information. The bookkeeping tables contain data set and run identification information used internally by the software. The flight-centric tables contain flight data such as flight plan information, track data, and controller information messages.

The concept of flight-centricity is key to the scenario generation process. During data extraction a flight is inserted into the database for each unique flight encountered in the recorded field data that had a flight plan message. Each of these extracted flights has a start time, which is the time of the flight's first recorded track message. All other events (e.g. flight plan messages, hold messages, interim altitude messages, and individual track messages) associated with the flight are maintained relative to this start time. Each flight also has a delta time, which is the amount of time that flight is to be shifted in time. During the data modification step an entire flight can be shifted in time but retain its flight profile merely by adding this delta time to the flight's start time.

Table 2.1-2: Air Traffic Database

Table Category	Table Name	Table Description
Environmental tables	*fd_airspace*	The *fd_airspace* table contains the constants required for x-y to lat-long coordinate conversion for a specific air traffic control facility.
	fd_rtix	The *fd_rtix* table contains the preferential routes names, indices, and types.
	fd_sector_asgn	The *fd_sector_asgn* contains information associated with the sector assignment messages.
	fd_route_status	The *fd_route_status* table contains information associated with the route status messages.
Bookkeeping tables	*fd_data_id*	The *fd_data_id* table contains information about a specific data extraction.
	fd_run	The *fd_run* table contains information identifying the data sets to be used for a specific run.
Flight-centric tables	*fd_flight*	The *fd_flight* table contains the static information related to a flight.
	fd_flight_plan	The *fd_flight_plan* table contains a history of the flight plans for a flight.
	fd_track	The *fd_track* table contains the individual track points for a flight.
	fd_int_alt	The *fd_int_alt* table contains a history of the interim altitude messages for a flight.
	fd_hold	The *fd_hold* table contains a history of the hold messages for a flight.
	fd_pref_route	The *fd_pref_route* table contains the converted route information for a flight.

2.1.3.1.2.2 Data Modification

The second step in the scenario generation process is data modification, which for this study consisted of time shifting the flights. This time shifting consisted of determining flight specific time increments that were added to the events associated with each of the flights. This caused each flight to follow it's recorded flight profile, but at a different time. As a result, aircraft-to-aircraft encounters and conflicts occurred in the scenarios that did not exist in the field.

The time shifting was accomplished using a genetic algorithm implemented in a program named *Cat*,[3] which was developed using:

- *gcc* Version 2.7.2.3, the GNU C/C++ compiler
- *libg+* Version 2.7.2, the GNU C/C++ libraries
- *Pro*C/C++* Version 8.1.6, the Oracle preprocessor that provides a software interface to tables within an Oracle Version 8.1.6 relational database

The goal of *Cat* is to find a set of delta times that can be applied to the flights in a scenario so that the distribution of parameters characterizing aircraft-to-aircraft conflicts meets user defined distribution constraints. A study determining the feasibility of using a genetic algorithm to time shift flights within a scenario is described in [Oaks, 2002]. A description of the implementation is presented in [Oaks and Paglione, 2002].

2.1.3.1.2.3 Scenario Generation

The final step in the scenario generation process is the actual generation of the scenarios using the time-shifted data. These scenarios are created in multiple formats, including the binary formatted files used by Lockheed Martin Air Traffic Management (LMATM) for URET Core Capability Limited Deployment (CCLD) accuracy testing and ASCII formatted files used by both LMATM and CPAT as input to other test tools. For the URET Prototype used in this study's experiment, the format generated is a MITRE ASCII defined format, called the SCN format. (See [Lindsay, 1997b] and [Lindsay, 1998].)

2.1.3.1.3 Summary of Genetic Algorithm Run

For this weather study *Cat Revision 1.57* was launched on a Sun Ultra 60 workstation with dual 450 MHz processors under the Solaris 8 operating system interfacing with an Oracle 8.1.6 relational database.

2.1.3.1.3.1 Input Parameters

The following summarizes the input to *Cat*:[4]

- The *DesiredFit* parameter was set to 0.99.
- The *SaveFit* parameter was set to 0.7.
- The *MinConflictCount* was set to 3.

Table 2.1-3 summarizes the constraint bin bounds.

- The *MaxGen* parameter was set to 2000.
- The *PopNbr* parameter was set to 30.
- The *Pc* parameter was set to 0.75.
- The *Pm* parameter was set to 0.01.
- The *Seed* parameter was set to 1.
- The *FitnessFlag* parameter was set to 2.
- The *SelectionFlag* parameter was set to 2.
- The *CrossoverFlag* parameter was set to 2.
- The *DistributionFlag* was set to u.
- The *geneLoBound* was set -1200.

---

[3] Cat was named for the character Cat on the British television series Red Dwarf. Cat is a humanized feline; the result of 3,000,000 years of evolution on the space ship Red Dwarf after all but one of its crew were killed by a radiation leak.

[4] A detailed description of these input parameters is provided in [Oaks and Paglione, 2002].

- The *geneHiBound* was set to 1200.
- The *Elitist* parameter was set to 4.

Table 2.1-3: Input Constraint Bin Bounds

Parameter	xx = Lo	xx = Hi
DesiredHorz1xx	55	67
DesiredHorz2xx	69	85
DesiredHorz3xx	46	56
DesiredHorz4xx	59	73
DesiredHorz5xx	18	22
DesiredVert1xx	184	224
DesiredVert2xx	37	45
DesiredVert3xx	9	11
DesiredVert4xx	18	22
DesiredVert5xx	0	2
DesiredAngl1xx	50	62
DesiredAngl2xx	37	45
DesiredAngl3xx	50	62
DesiredAngl4xx	55	67
DesiredAngl5xx	55	67
DesiredTypeVert1xx	64	78
DesiredTypeVert2xx	142	174
DesiredTypeVert3xx	41	51

2.1.3.1.3.2 Analysis of Run

Cat found a solution that met all of the constraints in 1043 generations. This took 10 hours 43 minutes 41 seconds. Figure 2.1-2 is a plot of fitness vs. generation number. The thicker line represents the value of the fitness of the best-fit chromosome in the generation and the thinner line represents the average fitness of the chromosomes in the generation (chromosome is the solution, see [Oaks and Paglione, 2002]). Figure 2.1-3 is a plot of the standard deviation vs. generation number. Figure 2.1-4 is a plot of the number of conflicts vs. generation number[5]. Table 2.1-4 shows the distribution of the conflicts in the constraint bins using the time shifts in this solution. It can be seen that all of the constraint counts fall within the desired low and high bounds. Figure 2.1-5 shows the distribution of the individual time shifts for the flights in the solution. The average of these time shifts for this solution was 25.687 seconds with a standard deviation of 674.521 seconds. The most a flight was shifted earlier in time was 1190 seconds; the most later in time was 1200 seconds.

[5] *Cat* estimates the number of conflicts to simplify processing. The CPAT conflict prediction accuracy tools calculated the actual conflict count used to evaluate URET in Section 3.2. It identified 211 conflicts, while *Cat* ended with 265 conflicts.

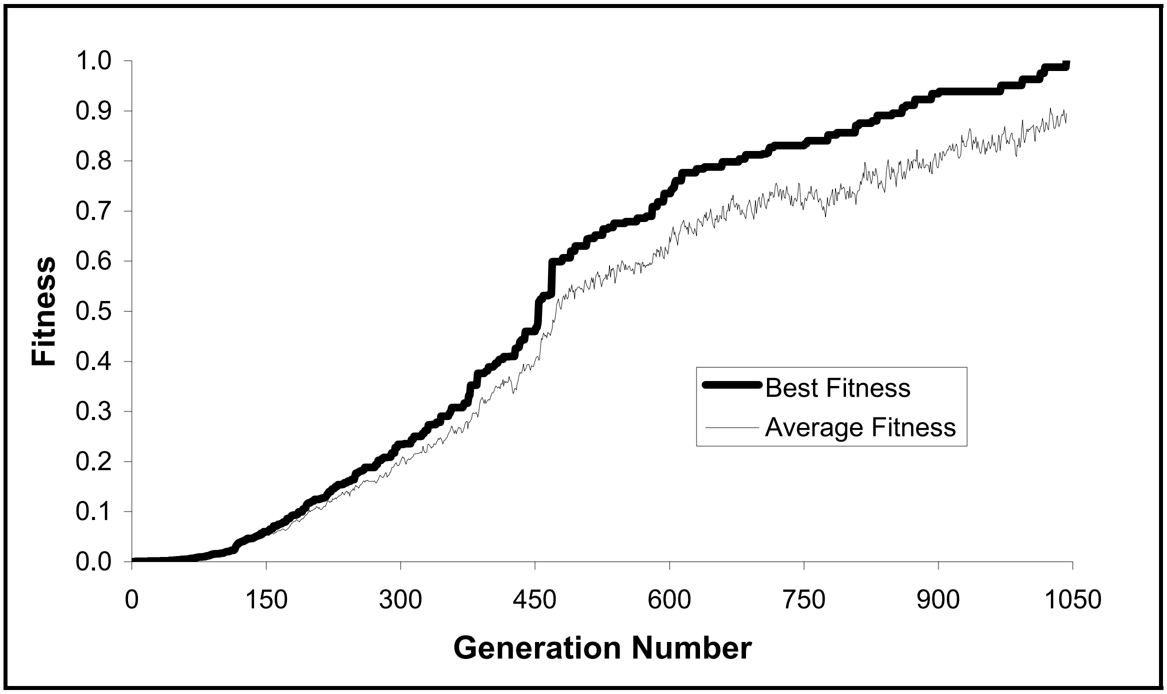


Figure 2.1-2: Fitness vs. Generation Number

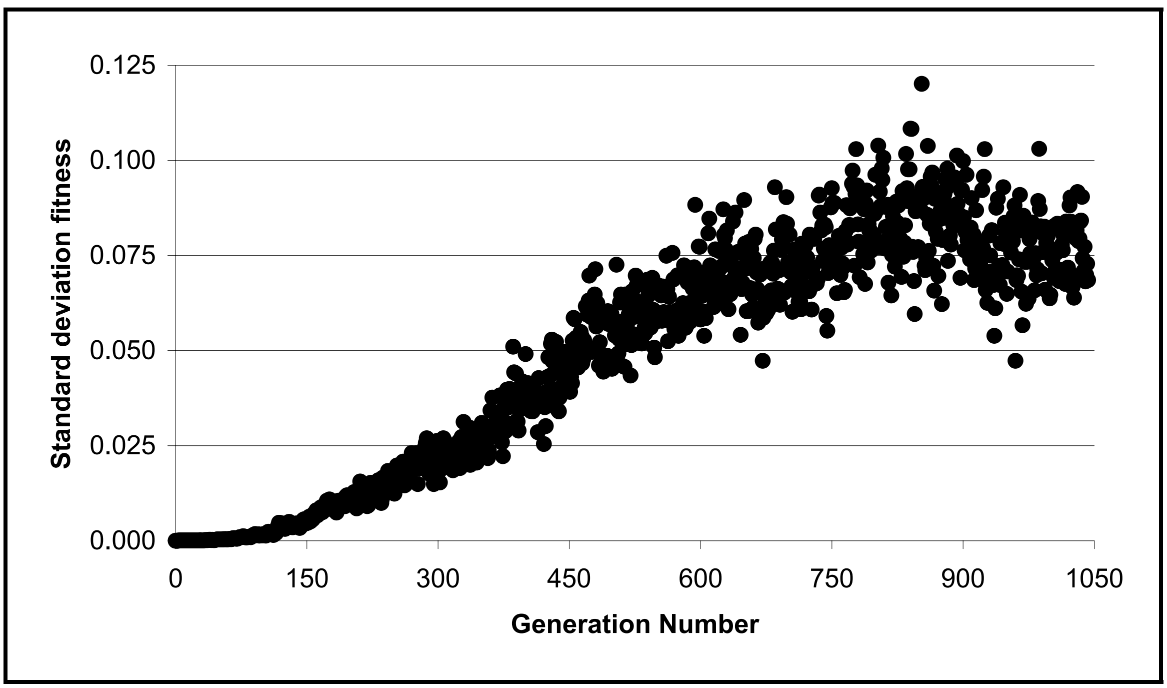


Figure 2.1-3: Standard Deviation Fitness vs. Generation Number

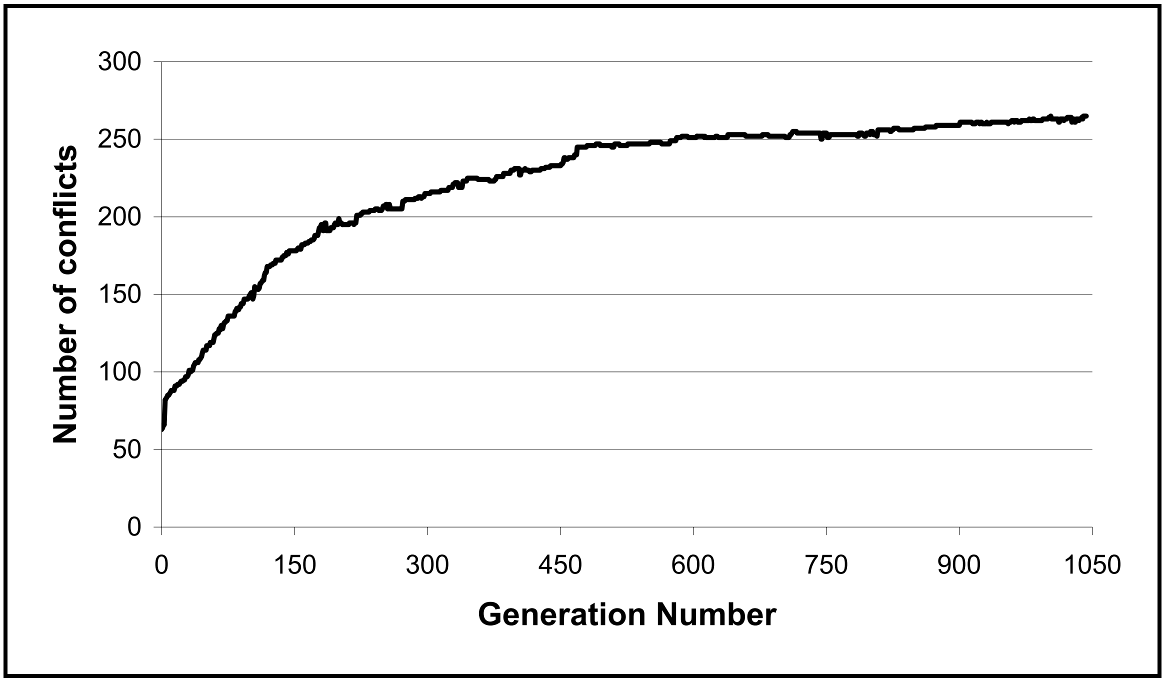


Figure 2.1-4: Number of Conflicts vs. Generation Number

Table 2.1-4: Constraint Bin Conflict Counts

Constraint	Low	Count	High
Horiz: 0 to 1 nm	55	56	67
Horiz: 1 to 2 nm	69	77	85
Horiz: 2 to 3 nm	46	52	56
Horiz: 3 to 4 nm	59	62	73
Horiz: 4 to 5 nm	18	18	22
Vert: 0 to 400’	184	195	224
Vert: 400 to 800’	37	40	45
Vert: 800 to 1200’	9	10	11
Vert: 1200 to 1600’	18	18	22
Vert: 1600 to 2000’	0	2	2
Angle: 0 to 36°	50	59	62
Angle: 36 to 72°	37	44	45
Angle: 72 to 108°	50	51	62
Angle: 108 to 144°	55	55	67
Angle: 144 to 180°	55	56	67
Level-level	64	68	78
Level-transitioning	142	146	174
Transitioning-transitioning	41	51	51

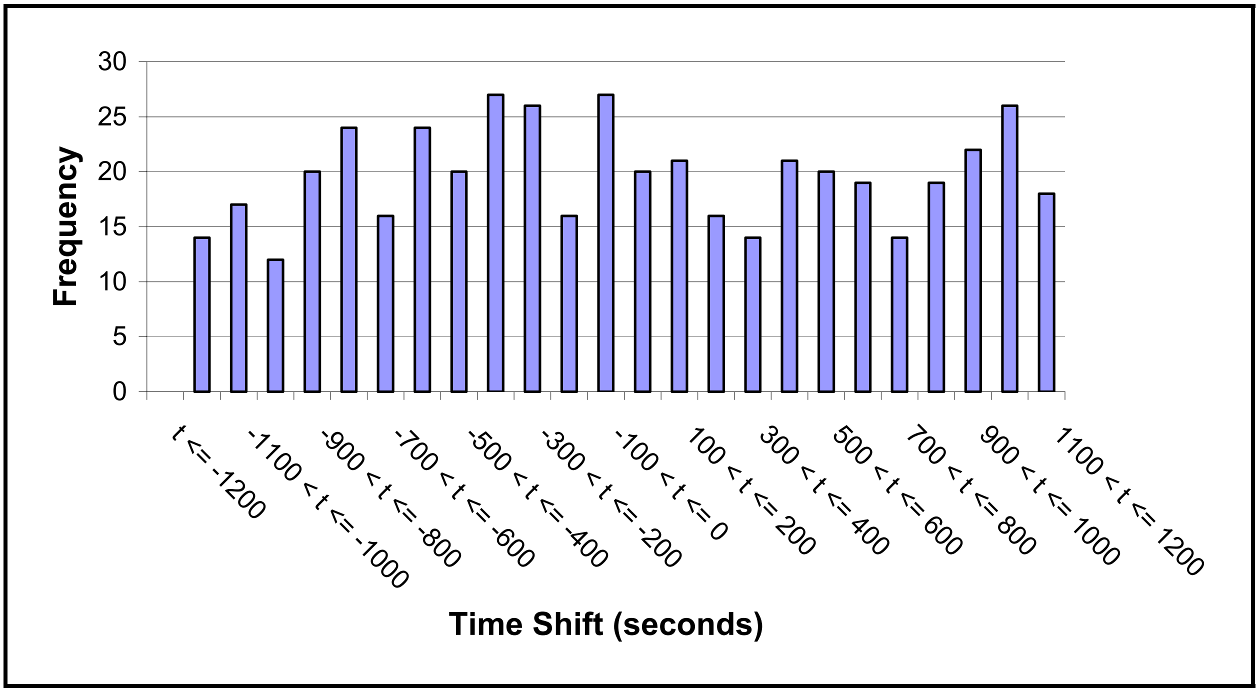


Figure 2.1-5: Distribution of Time Shifts

2.1.3.2 Input Weather Files

Forecasted weather data is provided to URET in Rapid Update Cycle (RUC) files using Advanced Weather Interactive Process System (AWIPS) Grid 211. These RUC files provide data for the continental United States (CONUS) using a Lambert conformal map projection that has a horizontal grid size of 80 kilometers on a side, which results in 93 points in the east-west direction and 65 points in the north-south direction. The forecasted wind magnitude, wind direction, and air temperature at 37 different altitude levels is provided at the nodes of this grid.

2.1.3.2.1 Description of RUC files

The data contained in these RUC files is formatted in the International Weather Meteorological Organization format known as gridded binary (GRIB), which is an international standard for the exchange of weather data (see [Johnson and Dey, 2000]). Table 2.1-5 (extracted from the GRIB format description [Johnson and Dey., 2000]) summarizes the RUC 211 grid.

Table 2.1-5: RUC 211 Grid Description

Projection	**Regional - CONUS (Lambert Conformal)**
Nx =	93
Ny =	65
La1 =	12.190N
Lo1 =	226.541E = 133.459W
Res. & Comp. flag =	0 0 0 0 1 0 0 0
Lov =	265.000E = 95.000W
Dx = Dy =	81.2705 km
Projection Flag =	0 (not bipolar)
Scanning Mode (Bits 1 2 3) =	0 1 0
Latin 1 =	25.000N
Latin 2 =	25.000N (tangent cone)
Grid Corners	(1, 1) = 12.190N, 133.459W (1, 65) = 54.536N, 152.856W (93, 65) = 57.290N, 49.385W (93, 1) = 14.335N, 65.091W
Pole is at (I, J) =	(53.000, 178.745)
The Dx, Dy grid increment (at 25 deg north) was selected so that the grid spacing at 35 deg north would be exactly =	80.000 km
The intersection of 35N & 95W falls on point =	(53, 25)

A specific set of RUC files contains an analysis file, which represents the weather conditions at a reference time, and forecast files, which provide the forecasted weather conditions. The availability of the RUC files and the number of forecast hours depends on the base data time. Table 2.1-6 summarizes this availability.

These RUC files include data for mean sea level, tropopause level, maximum wind level, freezing level, layers near the ground and surface, and 37 isobaric levels representing every 25 millibars (mb) between 1000 mb and 100 mb. Of this URET only uses the height (denoted as `HGT`) , east-west wind component (denoted as `UGRD`), the north-south wind component (denoted as `VGRD`), and the air temperature (denoted as `TMP`) at the 37 isobaric levels. The identification and location of this data in the RUC files is presented in Table 2.1-7 through Table 2.1-9. The URET trajectory modeler uses the barometric pressure (height) data in converting between Indicated Airspeed values and True Airspeed values, the air temperature data is used in converting between True Airspeed values and Mach numbers, and the wind vector is used to obtain the Ground Speed Vector from the aircraft airspeed vector.

Table 2.1-6: Availability of RUC 211 Files and Number of Forecast Hours

Base Data Hour	Approximate Availability	Analysis Hour	Forecast Hours (* indicates next day)
00:00	01:20	00:00	01:00 02:00 03:00 06:00 09:00 12:00
01:00	02:00	01:00	02:00 03:00 04:00
02:00	03:00	02:00	03:00 04:00 05:00
03:00	04:15	03:00	04:00 05:00 06:00 09:00 12:00 15:00
04:00	05:00	04:00	05:00 06:00 07:00
05:00	06:00	05:00	06:00 07:00 08:00
06:00	07:15	06:00	07:00 08:00 09:00 12:00 15:00 18:00
07:00	08:00	07:00	08:00 09:00 10:00
08:00	09:00	08:00	09:00 10:00 11:00
09:00	10:15	09:00	10:00 11:00 12:00 15:00 18:00 21:00
10:00	11:00	10:00	11:00 12:00 13:00
11:00	12:00	11:00	12:00 13:00 14:00
12:00	13:20	12:00	13:00 14:00 15:00 18:00 21:00 00:00*
13:00	14:00	13:00	14:00 15:00 16:00
14:00	15:00	14:00	15:00 16:00 17:00
15:00	16:15	15:00	16:00 17:00 18:00 21:00 00:00* 03:00*
16:00	17:00	16:00	17:00 18:00 19:00
17:00	18:00	17:00	18:00 19:00 20:00
18:00	19:15	18:00	19:00 20:00 21:00 00:00* 03:00* 06:00*
19:00	20:00	19:00	20:00 21:00 22:00
20:00	21:00	20:00	21:00 22:00 23:00
21:00	22:00	21:00	22:00 23:00 00:00* 03:00* 06:00* 09:00*
22:00	23:00	22:00	23:00 00:00* 01:00*
23:00	00:00	23:00	00:00* 01:00* 02:00*

Table 2.1-7: Geopotential Height Data in the RUC 211 Files (HGT)

Record Number	Type of Level Or Layer	Value	Data Units
1	level of 0 deg (C) isotherm	n/a	gpm
2	ground or water surface	n/a	
3	isobaric level	1000 mb	
4	isobaric level	975 mb	
5	isobaric level	950 mb	
6	isobaric level	925 mb	
7	isobaric level	900 mb	
8	isobaric level	875 mb	
9	isobaric level	850 mb	
10	isobaric level	825 mb	
11	isobaric level	800 mb	
12	isobaric level	775 mb	
13	isobaric level	750 mb	
14	isobaric level	725 mb	
15	isobaric level	700 mb	
16	isobaric level	675 mb	
17	isobaric level	650 mb	
18	isobaric level	625 mb	
19	isobaric level	600 mb	
20	isobaric level	575 mb	
21	isobaric level	550 mb	
22	isobaric level	525 mb	
23	isobaric level	500 mb	
24	isobaric level	475 mb	
25	isobaric level	450 mb	
26	isobaric level	425 mb	
27	isobaric level	400 mb	
28	isobaric level	375 mb	
29	isobaric level	350 mb	
30	isobaric level	325 mb	
31	isobaric level	300 mb	
32	isobaric level	275 mb	
33	isobaric level	250 mb	
34	isobaric level	225 mb	
35	isobaric level	200 mb	
36	isobaric level	175 mb	
37	isobaric level	150 mb	
38	isobaric level	125 mb	
39	isobaric level	100 mb	

Table 2.1-8: Temperature Data in the RUC 211 Files (TMP)

Record Number	Type of Level Or Layer	Value	Data Units
40	layer between two levels at specified pressure difference from ground to level	180-150 mb	K
41	layer between two levels at specified pressure difference from ground to level	90-60 mb	
42	layer between two levels at specified pressure difference from ground to level	30-0 mb	
43	specified height above ground	2 m	
44	isobaric level	1000 mb	
45	isobaric level	975 mb	
46	isobaric level	950 mb	
47	isobaric level	925 mb	
48	isobaric level	900 mb	
49	isobaric level	875 mb	
50	isobaric level	850 mb	
51	isobaric level	825 mb	
52	isobaric level	800 mb	
53	isobaric level	775 mb	
54	isobaric level	750 mb	
55	isobaric level	725 mb	
56	isobaric level	700 mb	
57	isobaric level	675 mb	
58	isobaric level	650 mb	
59	isobaric level	625 mb	
60	isobaric level	600 mb	
61	isobaric level	575 mb	
62	isobaric level	550 mb	
63	isobaric level	525 mb	
64	isobaric level	500 mb	
65	isobaric level	475 mb	
66	isobaric level	450 mb	
67	isobaric level	425 mb	
68	isobaric level	400 mb	
69	isobaric level	375 mb	
70	isobaric level	350 mb	
71	isobaric level	325 mb	
72	isobaric level	300 mb	
73	isobaric level	275 mb	
74	isobaric level	250 mb	
75	isobaric level	225 mb	
76	isobaric level	200 mb	
77	isobaric level	175 mb	
78	isobaric level	150 mb	
79	isobaric level	125 mb	
80	isobaric level	100 mb	

Table 2.1-9: North-South Component of the Wind Data in the RUC 211 Files (VGRD)

Record Number	Type of Level Or Layer	Value	Data Units
124	layer between two levels at specified pressure difference from ground to level	180-150 mb	m/s
125	layer between two levels at specified pressure difference from ground to level	90-60 mb	
126	layer between two levels at specified pressure difference from ground to level	30-0 mb	
127	specified height above ground	10 m	
128	Maximum wind level	n/a	
129	tropopause	n/a	
130	isobaric level	1000 mb	
131	isobaric level	975 mb	
132	isobaric level	950 mb	
133	isobaric level	925 mb	
134	isobaric level	900 mb	
135	isobaric level	875 mb	
136	isobaric level	850 mb	
137	isobaric level	825 mb	
138	isobaric level	800 mb	
139	isobaric level	775 mb	
140	isobaric level	750 mb	
141	isobaric level	725 mb	
142	isobaric level	700 mb	
143	isobaric level	675 mb	
144	isobaric level	650 mb	
145	isobaric level	625 mb	
146	isobaric level	600 mb	
147	isobaric level	575 mb	
148	isobaric level	550 mb	
149	isobaric level	525 mb	
150	isobaric level	500 mb	
151	isobaric level	475 mb	
152	isobaric level	450 mb	
153	isobaric level	425 mb	
154	isobaric level	400 mb	
155	isobaric level	375 mb	
156	isobaric level	350 mb	
157	isobaric level	325 mb	
158	isobaric level	300 mb	
159	isobaric level	275 mb	
160	isobaric level	250 mb	
161	isobaric level	225 mb	
162	isobaric level	200 mb	
163	isobaric level	175 mb	
164	isobaric level	150 mb	
165	isobaric level	125 mb	
166	isobaric level	100 mb	

2.1.3.2.2 Description of Software Tools Used to Process Weather Data

The following subsections describe the software acquired from the National Weather Service (NWS) and the software developed by CPAT.

2.1.3.2.2.1 Software Developed by the National Weather Service

This section describes the shareware tools from the NWS utilized in this study.

2.1.3.2.2.1.1 wgrib

wgrib was written by the NWS and is available to the public on the Internet (see http://wesley.wwb.noaa.gov/wgrib.html). Its primary use is to generate subsets of the RUC GRIB data in a format (text, binary, IEEE, etc) specified by the user. The raw RUC files comprising this data may contain more than 225 weather parameters. A typical use of this program is to generate a high-level summary for some weather parameter such as `u-wind` data at 1000 millibars (mb). To generate the data for a given parameter the user must know its position within the RUC file. This program does not support filtering on a geographic basis and performs no analysis functions of a GRIB record other than specifying its minimum and maximum values. The output file named by the user contains a single value per line in the user specified format.

2.1.3.2.2.1.2 gribw

gribw was written by the NWS and a beta version is available to the public on the Internet (see http://wesley.wwb.noaa.gov/gribw.html). The program functions as the inverse of *wgrib* in that it can create a GRIB-formatted file from a binary file and a text file containing the Product Description Section (PDS) and Grid Description Section (GDS) created by *wgrib*.

2.1.3.2.2.2 Software Developed by CPAT

This section describes the weather processing software tools developed by the CPAT utilized in this study.

2.1.3.2.2.2.1 getRUC

getRUC populates an Oracle database weather table with data extracted from a binary file generated by *wgrib*. This table is described in Table 2.1-10. The naming convention for this table is wx*mmddyyyy*h*hh*f*ff* where:

mm	is the month
dd	is the day
yyyy	is the year
hh	is the base data hour for the RUC data
ff	is the forecast hour for the RUC data

For example wx05261999h09f01 contains the forecasted weather data for 10:00 for May 26, 1999 (i.e., base data hour = 09:00 and forecast hour = 1 hour).

Table 2.1-10: Description of the Weather Table

Field Name	Description	Units	Data Type
ruc_i	node number along a latitude circle	n/a	number(2)
ruc_j	node number along a longitude meridian	n/a	number(2)
ruc_hgt	geopotential height	gpm	number(5)
ruc_tmp	Temperature	degrees Kelvin	number(4,1)
ruc_ugrd	u component of wind	meters/second	number(5,1)
ruc_vgrd	v component of wind	meters/second	number(5,1)
lat	latitude of node	degrees	number(7,4)
lon	longitude of node	degress	number(7,4)
x	x-coordinate	nautical miles	number(7,3)
y	y-coordinate	nautical miles	number(7,3)
altitude	Altitude	feet	number(6,1)
air_temp	air temperature	degrees Celsius	number(4,1)
wind_mag	magnitude of wind vector	knots	number(5,1)
wind_dir	direction of wind vector	degrees	number(5,1)

2.1.3.2.2.2.2 modRUC

modRUC processes a binary file created by *wgrib*, producing another binary file containing modified values reflecting changes to the temperature and to the magnitude and direction of the wind vector.

Within the RUC weather files, the air temperature is represented in degrees Kelvin. *modRUC* modifies all the air temperatures in the RUC file by adding an input value measured in degrees Celsius, which is equivalent to degrees change in Kelvin since one degree Kelvin is equivalent to one degree Celsius.

modRUC modifies the magnitude of the wind vector by adding an input value measured in knots and modifies the direction of the wind vector by rotating the vector clockwise through an input angle measured in degrees. This modification is shown in Figure 2.1-6. The wind vector (v_w) has a magnitude of $|v_w|$ and a direction of θ. Changing its magnitude by the value Δv_w and rotating it through the angle $\Delta\theta$ results in the vector v_w', where

$$v_w' = (\textit{u-wind}', \textit{v-wind}')^T$$

where

$$u\_wind' = (|v_w| + |\Delta v_w|) \sin(\theta + \Delta\theta)$$
$$v\_wind' = (|v_w| + |\Delta v_w|) \cos(\theta + \Delta\theta)$$

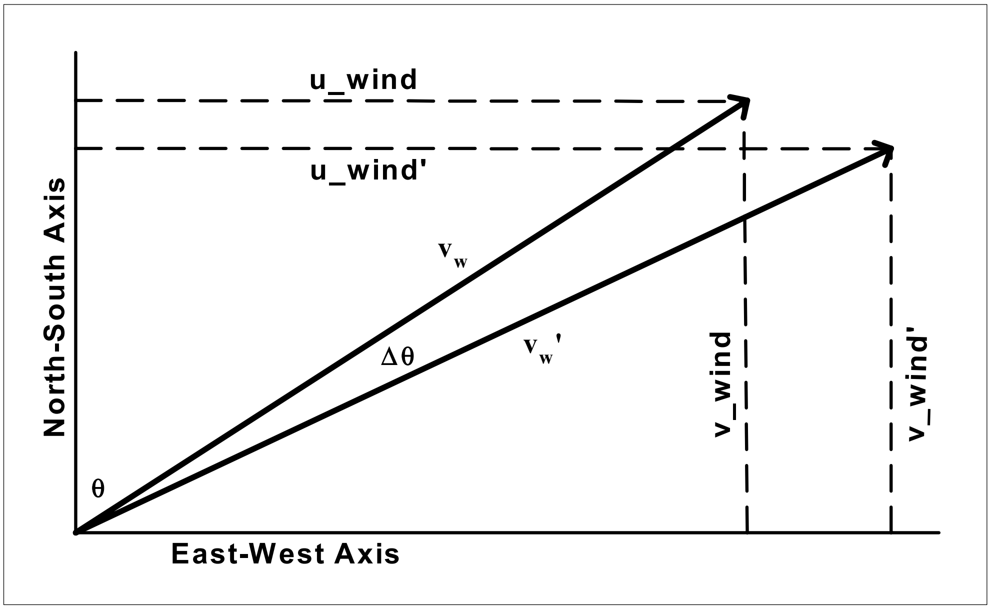


Figure 2.1-6: Modification of Wind Vector

2.1.4 Output Data of Experiment

For this study, a special release of the MITRE CAASD URET Prototype was used, referred to as URETD32R2P1C Lockheed Martin release, and was run in the Free Flight Technology Integration Laboratory (FFTIL). This URET Prototype was used to test the production version of URET, URET Core Capability Limited Deployment (CCLD), in the FAA Formal Testing Program. This version of the URET Prototype was utilized in this weather study, since its accuracy closely matched that of URET CCLD and was easily accessible to CPAT. Furthermore, the URET Prototype has an extensive data recording capability. For this study, three output records were used, which archive:

1. All trajectories (referred to as record 16 in the URET software)
2. All route conversions (referred to as record 15)
3. All conflict notifications or alerts (referred to as record 82)

The details of extracting these records from the URET Prototype software are documented in [Jaarsma et al., 1997] and URET's data flow of these data structures in [Bashioum and Mayo, 1997]. CPAT's processing of these URET output records are documented thoroughly in [Cale et al., 1998] and [Paglione et al., 1999a]. Briefly, CPAT parsed the trajectory records or as referred to by the URET software the State Segment (SSG) data structures. These trajectory records were sampled by CPAT to perform the trajectory accuracy analysis discussed in Section 2.2.2.

CPAT also parsed the converted routes URET produced. These route segments are referred to as Onboard Route Segments (ORS) by the URET software. The ORS records provided a listing of the horizontal path coordinates expanded from the HCS flight plan and/or amendment messages. These records were used by CPAT to determine if an aircraft was within certain thresholds laterally of its cleared route of flight. This is referred to as flight plan adherence and will be discussed in Section 2.2.3.

CPAT processed the URET alert records next. These URET output records contain the notification times, predicted start and end times, predicted minimum horizontal separation, alert types (e.g. red, yellow, muted), and other related fields for each conflict notification action URET executed. Each record contained an action field that identified whether the alert was just added to the URET display for the first time, modified in some way, or deleted. These records were the main source used to analyze the conflict prediction accuracy of URET. The details of this analysis are presented later in Section 2.2.3.

2.2 Description of Analysis

This section will present the analysis methodology of the study. First, Section 2.2.1 briefly presents the analysis involved in the parsing and the processing of the HCS messages, which drive the experiment. Next, Section 2.2.2 presents the trajectory accuracy methodology, and finally Section 2.2.3 describes the methodology for analyzing the resulting URET conflict predictions.

2.2.1 Scenario Processing

The air traffic scenario used in this study consists of aircraft surveillance reports (i.e. track reports) and HCS clearances as discussed in Section 2.1.3.1. Besides being used as the main input into the URET system, the messages form the basis of the accuracy analysis. The track reports provide a four-dimensional set of positions, defining the actual path the aircraft flew. These reports are parsed, checked for reasonableness, and interpolated for later accuracy measurements against the URET predictions. The track reports provide the aircraft positions against which the trajectory predictions are evaluated. They are also used to calculate aircraft-to-aircraft separations, resulting in a database of conflicts (i.e. violations of ATC standard separation distances) and encounters (i.e. violations of specified separation distances). These conflict and encounter events are then used to compute the accuracy of URET's trajectory and conflict predictions.

The most common HCS clearances include flight plan messages and interim altitude messages. To build trajectories, URET uses the flight plan messages for the aircraft planned routes and interim altitude messages as temporary hold altitudes. These messages are parsed to allow later statistics to be performed, such as counts on type of aircraft, equipage, and flight type (i.e. over flight, arrival, departure, and internal). The details of these processing steps were first presented in [Cale et al., 1998] and later in [Paglione et al., 1999a].

2.2.2 Trajectory Accuracy Measurement

Trajectory accuracy measurement is a process of sampling trajectory positions using HCS track position reports and calculating the time coincident trajectory errors in the horizontal and vertical dimensions. The sampling process, named the interval based sampling technique, is described in detail first in [Paglione et al., 1999a] and later in [Cale et al., 2001]. Briefly, the interval based sampling technique is a two-step process that pairs the track and trajectory points to measure the prediction errors for an entire flight. This sampling technique takes the perspective of the DST user, the air traffic controller. The trajectory active at the sampling time is used for measurement. The predictions of the aircraft position at the sampling time, at the sampling time plus 300 seconds into the future, plus 600 seconds, etc. are recorded and compared to the actual position of the aircraft at those times. These incremental times are referred to as the look-ahead times.

For this study, the trajectory accuracy, utilizing the interval based sampling technique, is applied to each run of the experiment. The errors for the individual runs are then compared per measurement time, since all the runs use the same input scenario (i.e. track reports).

2.2.2.1 Trajectory Error Deviations Between Runs

For the trajectory accuracy analysis, the difference is calculated between the unsigned trajectory deviation error from one run minus the time coincident error measurement in a second run. Most of the analysis is the comparison of the control run (i.e. Run 000, see Table 2.1-1) minus a treatment run (e.g. 200, 100, 010, etc.), which indicates whether the induced error in the RUC variable influenced the trajectory performance. For example, the horizontal error comparison for the Run 000 to Run 100 (the low-level wind magnitude run) was generated as the measured horizontal error of Run 000 minus the corresponding time coincident error of Run 100. The trajectory deviations analyzed were horizontal error, the orthogonal components of horizontal error, unsigned lateral and longitudinal error, and the unsigned vertical error. The data was also filtered to exclude trajectory measurements made beyond outbound ARTCC hand-off and beyond any air traffic control directives. The filtering process is explained in detail in [Cale et al., 2001].

2.2.2.2 Trajectory Accuracy Statistical Methods

This section reports on statistical methods used in analyzing the trajectory error deviation between runs. The techniques include parametric and nonparametric measures and various univariate statistics provided in a standard statistical analysis package. A brief overview of this section is as follows.

First, a preliminary analysis of the trajectory error data shows the data to be non-Gaussian in distribution. This determination suggests the use of a nonparametric approach to the statistical analysis of the data. Next, background information on the designed experiment is provided. The experiment involved selecting factor levels, running the experiment with one or more levels set, and using hypothesis testing to determine if the results indicate a statistically significant effect due to the factor level. The Wilcoxon signed-rank test is next introduced. This is a nonparametric hypothesis testing method used when the data is determined to be non-Gaussian. Next, the SAS (Statistical Analysis System)[6] Proc Univariate procedure is introduced, which provides extensive statistics including results from the Wilcoxon signed-rank test. The final section provides general notes on the specific approach used to analyze the data.

2.2.2.2.1 Preliminary Data Analysis

For this study, the difference in trajectory errors between runs was found to be non-Gaussian in distribution, as shown in a set of Q-Q plots (quantile-quantile). The Q-Q plots represent the difference in the distributions of trajectory errors between the two runs. The Q-Q plot is a statistical method used to test whether a data sample can be considered as having been drawn from a larger population that is normally distributed. A determination of non-Gaussian limits the type of statistical test procedures that can be done. Figure 2.2-1 presents the Q-Q plots for horizontal and vertical trajectory errors.

The left-side of the figure is a Q-Q plot of the delta (control *minus* altered) in horizontal error for wind magnitude. The right-side of the figure shows the delta in vertical error for the same scenario and RUC files. Both plots show curvature in the tail regions indicating that the data is non-Gaussian. The quantiles of a sample drawn from a true Gaussian distribution will plot linearly when compared with quantiles of the proposed theoretical normal distribution. From these plots, we can conclude that the trajectory data for this sample is non-normal. Assuming this sample to be representative of all such samples, a nonparametric approach to the statistical analysis will be required.

[6] SAS is a popular statistical analysis software package utilized in this study.

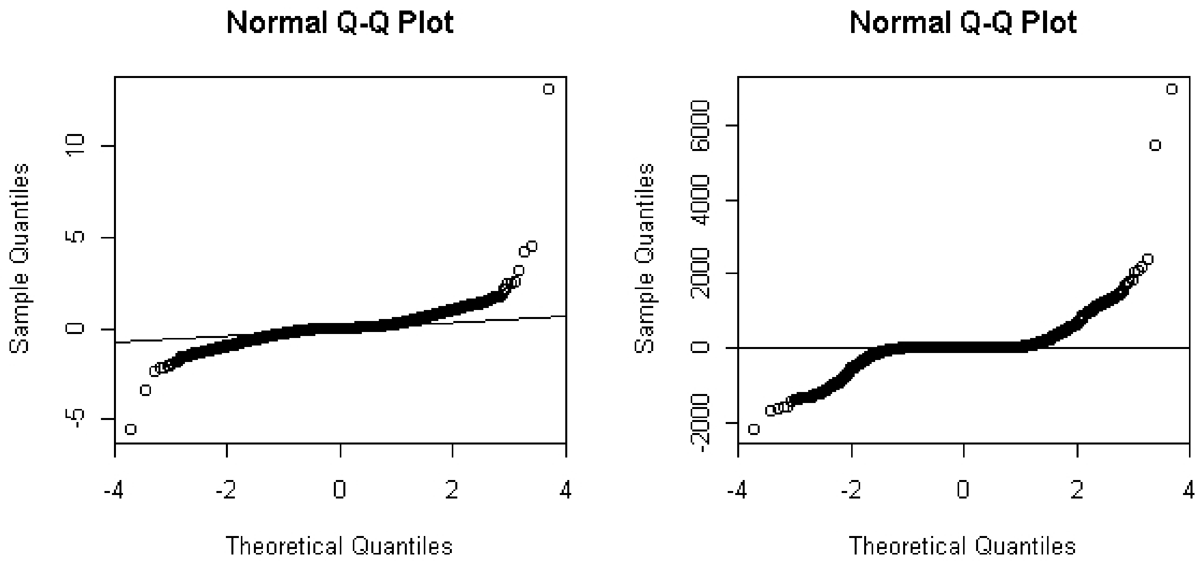


Figure 2.2-1: Q-Q Plots of Horizontal and Vertical Trajectory Deviation Error

2.2.2.2.2 Design of Experiment Methodology

A *designed experiment* is a statistical method used to identify which system input factors and which levels within the factors significantly affect the output response. The methodology is to first identify factors and levels that cover some meaningful range of input variables, next make multiple runs of the system at the identified factors and levels and finally statistically analyze the system response. In this study, the levels believed to cover a valid range for each weather forecast factor (i.e. the wind magnitude, wind direction, and air temperature as described in Section 2.1.2) were determined and multiple runs of the URET system were conducted using RUC files where one or more of the factors were altered. Next the results (measured trajectory deviation error) from the multiple runs were compared to determine which weather forecast factors had an impact on trajectory accuracy. With this approach it is also possible to alter multiple factors, run the experiment and get information on possible interactions between factors and factor levels. The study initially envisioned two blocks of runs to evaluate both single and combinations of multiple factors.

The focus of Run Set A was on the main weather forecast factors only (wind magnitude and direction, air temperature). This was accomplished by systematically altering one factor while holding the remaining two factors at their nominal level and running the experiment. The statistical analysis then compares the difference in trajectory error between a run using the altered file and a baseline run using the nominal file or between two altered files where the factors are at different predetermined levels. This type of analysis where only a single factor is altered provides information on which levels within that factor cause a statistically valid change (if any) in trajectory deviation error. Thus, the main effect of the weather forecast factors would be determined.

The intent of Run Set B was to provide information on possible interactions between the weather forecast factors. This would be accomplished by systematically altering the levels of two or more different factors, running the traffic scenario with the altered RUC file through the system and doing the trajectory analysis. This type of experiment would identify a change in trajectory accuracy caused by some combined interaction between the multiple weather factors. For example, perhaps some level of wind magnitude combined with some level of air temperature might cause an increase in trajectory error not observed when only one of these factors is altered.

The designed experiment is considered the traditional approach to statistical analysis of a system under study. Unfortunately the methods used to analyze the results (e.g. ANOVA, paired T-test) assume both normality in the distribution of sample data and also equal variance. As indicated in Section 2.2.2.2.1, the sampled trajectory deviation data was shown to be non-Gaussian in distribution and requires an alternative approach. This is described in the next Section 2.2.2.2.3.

2.2.2.2.3 Nonparametric Statistical Experiment

Since the trajectory sample data is non-normal in distribution, a nonparametric approach to statistical testing was utilized. Another technique considered was to transform the data to achieve normality and do an analysis on the new data set. This idea was discarded as the study involves the management of multiple data sets resulting from the numerous combinations of look-ahead times and trajectory error types. Also, correcting for normality may unequally affect sample variance in different data sets and equality of variance is a requirement for many of the traditional statistical tests.

Nonparametric data analysis is an alternative that makes few assumptions regarding the characteristics of the parent population. These methods are based on data ranks and use the corresponding sampling distributions to determine significance. This field of statistics is considered classical in the sense that the methods are well investigated, understood and accepted. The nonparametric test appropriate for this study is a pairwise comparison method called the Wilcoxon signed-rank test [Hollander and Wolfe, 1999].

The signed-rank test is used with *pre*- and *post*-treatment type observations where the interest is in determining whether the treatment or the experiment has caused a shift in the median value of the data. A shift would indicate that results from the post-treatment test were sampled from a second population not simply a second sampling of the original population. The procedure is to sort (ascending) and rank (sequentially) the absolute value of the differences between the two runs. The difference will be either positive or negative depending on which observation in the comparison was larger. Next the ranks are summed where the original test difference was positive. This sum is then compared either to a table value (small sample size) or to the standard normal distribution (large sample size) to decide if the positive sum of ranks is beyond some limit considered possible if the data were in fact from the same population. The theory is if the original before and after test results differ by a lot and in one direction (mostly negative or positive) then the summed positive (or negative) ranks will be either very large or very small. Table values or the standard normal distribution provide cut-off points for what is probabilistically too large or too small for the samples to come from a single population.

Two unforeseen problems were discovered with the signed-rank test: the exclusion of zero values in the test procedure and the lack of a nonparametric method for testing interaction effects between main effects. Both of these will be briefly discussed next.

Zero values in the data occur when the treatment application had literally no effect and the differences (i.e. control *minus* treatment) equal zero. The test excludes these values and the analysis is based on the remaining observations. This appears to be a flaw in the test as a zero result in a paired replicates analysis would strongly support a hypothesis of no difference between treatments. An example of this situation and the solution utilized in this study are provided in Section 2.2.2.2.6.

Run Set B of the study was to investigate interactions between the main factors. A nonparametric method for testing interactions is not available in the SAS statistical package. A literature survey provided a single paper on nonparametric methods to test interactions in experimental [Sawilowsky, 1990]. As a consequence, this study focused on the testing of main factors only.

2.2.2.2.4 SAS Univariate Statistical Procedure

The study made extensive use of the SAS Proc Univariate (single factor or variable) procedure. A brief summary of the statistical measures provided in the SAS report are as follows: data moments (number of observations, sample mean, median, mode), variability (standard deviation, variance, range, inter-quartile range), moments (skewness and kurtois), tests for location (Student-t, Wilcoxon sign and signed-rank tests), several tests for normality, the quantiles and extreme observations. The signed-rank test was the primary indicator of the test of statistical significance, but extensive use was made of the inter-quartile range and quantiles where the data had many zeros or did not have *practical significance* [Devore, 2000]. This last term is defined in Section 2.2.2.2.6.

2.2.2.2.5 Graphical Statistical Plots

The SAS JMP statistical package was used to produce quantile box plots. These quantile box plots were made for each error type partitioned by look-ahead time. The plot graphically depicts the same information available from the quantiles. Specifically, the horizontal lines in the quantile box plot represent (from top to bottom) the 95^{th}, 75^{th}, 50^{th}, 25^{th}, and 10^{th} quantiles or percentiles. These components are graphically presented in Figure 2.2-2. An example application of one of these SAS JMP box plots is illustrated in Figure 2.2-3. In this example, the spread increases and the median decreases as look-ahead time increases. The full range of SAS JMP box plots are presented in Appendix A.

Microsoft Excel was used for each run comparison to plot the median value at each look-ahead time and error type. Figure 2.2-4 provides an example plot showing the median is decreasing (becoming more negative) as the look-ahead time increases. Thus, for this example plot, the horizontal error for each run’s measurements is larger for the treatment run versus the control run. The full range of these median plots are presented in Appendix B.

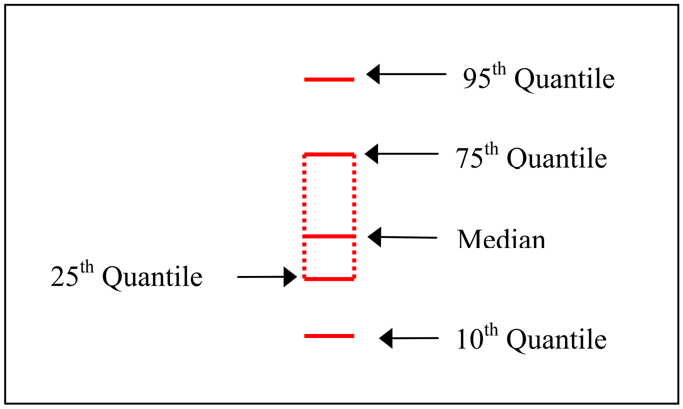


Figure 2.2-2: Components of JMP Quantile Box Plot

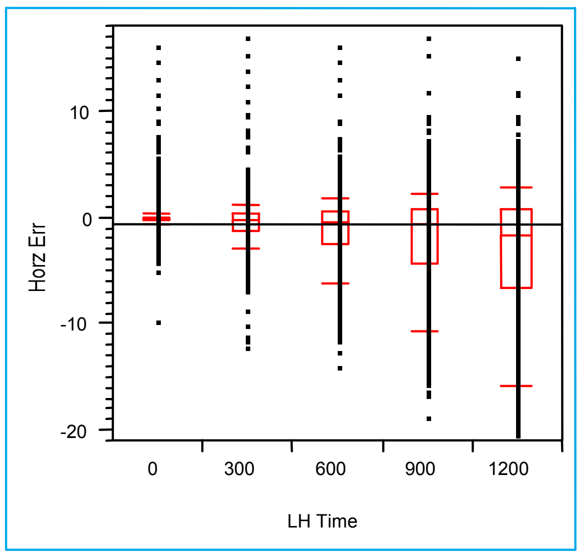


Figure 2.2-3: Example JMP Box Plot of Horizontal Errors for Comparison 000-200

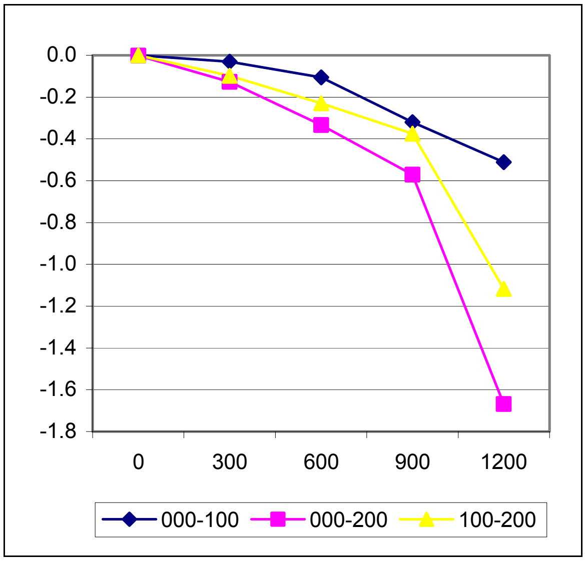


Figure 2.2-4: Plot of Median Difference in Horizontal Error by Look-Ahead Time for Wind Magnitude during Level Flight

2.2.2.2.6 Applied Statistical Analysis

This section provides detailed notes of the statistics applied in Section 3.1, which presents the results of the study. The horizontal and vertical phase-of-flight classifications, the problem with insufficient data for two of the phase-of-flight groupings, and the use of practical significance in determining statistical significance are discussed in this section.

The raw trajectory deviation data has three vertical and three horizontal phase-of-flight classifications. These classifications were consolidated to reduce the overall size of the analysis. The original horizontal classifications of straight (STR), left turn (LFT) and right turn (RHT) were reduced to straight (STR) or turn (TRN). The original vertical classifications of level (LEV), descending (DES), and ascending (ASC) flight were reduced to level (LEV) or in-transition (TRAN) flight. The consolidation left four phase-of-flight groups – straight and level flight (STR-LEV), straight and in-transition (STR-TRAN), turn and level flight (TRN-LEV), and turn and in-transition (TRN-TRAN).

Statistical analysis was ultimately done on only two of these classifications. Sorting the data by consolidated phase-of-flight classification and then by look-ahead time left some combinations with insufficient observations for a valid statistical analysis. Table 2.2-1 provides the number of observations available by group and look-ahead time combination and the percentage of observations retained after filtering points beyond the center boundary or beyond an air traffic control directive. The table shows relatively few or even no observations for the TRN-LEV (turn and level) and TRN-TRAN (turn and in-transition) groups. Even within the STR-TRAN (straight and in-transition) group there are few observations beyond the 300 second look-ahead time. Therefore, the analysis was further limited to the STR-LEV (straight and level) group and to the zero and 300 look-ahead times for the STR-TRAN (straight and in-transition) group. In the remainder of the report the straight reference will be dropped and the two groups will be referred to as level (LEV) and in-transition (TRAN) flight.

Table 2.2-1: Count and Percentage of Observations Retained by Phase-of-Flight and Look-Ahead Time

Group	Look-Ahead Time (seconds)									
	0		300		600		900		1200	
STR-LEV	3056	100%	2102	82%	1301	64%	822	53%	488	45%
STR-TRAN	2041	100	207	15	22	2	2	<1	2	<1
TRN-LEV	57	100	33	65	15	43	8	38	5	38
TRN-TRAN	107	100	6	11	1	4	0	0	0	0

Another issue required validating the signed-rank test where the data contained an extensive number of zero observations. The study data are derived as the difference in trajectory deviation between a control run and a treatment run. A zero observation occurs where the experiment had no effect on trajectory deviation beyond what was observed in the control run. The signed-rank test eliminates zero observations and the analysis is based on the remaining observations only. This is a limitation in the technique as logically a zero observation provides strong evidence supporting a null hypothesis of no difference in run results. The phenomena became apparent when reviewing the lateral error portion of the SAS generated reports. For example, the quantiles for the 000-100 level flight run showed extensive zero observations, but the signed-rank test rejected the null hypothesis of no difference between runs. Table 2.2-2 provides the reported lateral error quantiles for the 000-100 run sequence and Table 2.2-3 provides the reported test results. The first table indicates that 90 percent of the data is either equal to or essentially equal to zero, yet the P-values[7] shown in the second table support a conclusion that the runs are statistically different.

Table 2.2-2: Quantiles for Run Set 000-100, Lateral Error and Level Flight

Quantile	Lateral Error
Max	1.1213nm
99%	0.2876
95	0.0083
90	0.0001
75	0.0000
50	0.0000
25	0.0000
10	-0.0011
5	-0.0341
1	-0.2814
Min	-2.0491

[7] In [Devore, 2000], the P-value is defined as the "smallest level of significance at which the null hypothesis would be rejected when a specified test procedure is used on a given data set." Thus, the P-value is the probability of the null hypothesis has occurred, so a small P-value (less than 0.10) would indicate the null hypothesis unlikely and should be rejected and if large should be assumed correct.

Table 2.2-3: Test Results for Run Set 000-100, Lateral Error and Level Flight

Test	P-value
Student-t	0.0661
Sign	0.0102
Signed-Rank	0.0014

Research into the signed-rank test revealed that the method is actually a test of symmetry with a large asymmetric result supporting a conclusion that the samples are actually drawn from two different populations. Data having zero values are excluded as they provide no information on symmetry. One reference indicated that a conservative approach for data with many zero observations is to simply count these as supporting the null hypothesis when assessing the test results [Hollander and Wolfe,1999]. This approach was applied extensively in validating test results for both lateral and vertical error.

Another issue assessed validity of the test results when the range of the data was small in magnitude and logically equal to zero. This phenomenon is known as *practical significance* [Devore, 2000]. Here a statistical significance test rejects the null hypothesis of no difference, but the scale of the data is such that a logical conclusion would support the null. For example, Table 2.2-4 provides quantiles for the 000-100, in-transition run set for vertical error. The data shows at least 50 percent of the error deviations to be within 100 feet. It is known that HCS reported aircraft positions only have a precision of 100 feet or more vertically. Therefore, any deviations within 100 feet are effectively zero.

Table 2.2-4: Quantiles for Vertical Error Run Set 000-100 and In-Transition Flight

Quantile	Vertical Error
Max	1316.35 ft
99%	931.0630
95	385.6491
90	128.0028
75	12.0192
50	0.0000
25	-20.0000
10	-150.6942
5	-389.5242
1	-982.2329
Min	-2183.7945

Table 2.2-5 provides the corresponding test results for the 000-100 in-transition run set. Here the signed-rank test loosely supports the null hypothesis (for an alpha rejection level=0.10 percent).

Table 2.2-5: Test Results for Run Set 000-100, Vertical Error and In-Transition Flight

Test	P-value
Student-t	0.4119
Sign	0.2777
Signed-Rank	0.1295

The lack of practical significance was also observed for horizontal, longitudinal, and lateral error where some run sets showed a large portion of the error data to be plus or minus a few tenths of a nautical mile about a zero median value. Essentially it was necessary to consider multiple statistics provided in the SAS reports (median, interquartile range, quantiles) to verify the statistical significance of each hypothesis test.

2.2.3 Conflict Prediction Accuracy Measurement

The objective of the analysis is to determine the difference of the aircraft-to-aircraft conflict predictions between the control run and the treatment runs. This difference establishes the conflict prediction impact of the induced weather forecast error on the treatment runs. The analysis can include predictions for aircraft-to-aircraft encounters or conflicts[8] and excludes errors not adhering to known air traffic control clearances (i.e. flight plan adherence) or ignores this characteristic. Flight plan adherence is defined in detail in [Paglione et al., 2000] and [Paglione and Summerill, 2000]. For this study, four analyses were performed for each URET run and include:

- Analysis AA - adherence included and red alerts only
- Analysis BA - adherence ignored and red alerts only
- Analysis AB - adherence included and both red and yellow alerts
- Analysis BB - adherence ignored and both red and yellow alerts

Adherence requires aircraft to be flying within specified thresholds laterally and vertically to the current known HCS clearance. If at the start time of the actual conflict (or predicted start time for potential false alerts) either one of the flights has not been in adherence for at least 13 minutes, then the conflict prediction will be discarded if not correctly notified. Regardless of the age of adherence at the start of the conflict, adherence only potentially excludes missed or false alerts. Thus, a correct prediction defined as a valid alert is not affected by adherence. For the Analysis AA and AB, including adherence discards some missed and false alerts, so presumably the accuracy results would be better than the analyses ignoring adherence. This is no surprise. When a conflict probe has improved flight intent input information, the predictions it makes are expected to be more accurate. By performing the analyses with and without adherence, the

[8] Conflicts between aircraft are in regard to standard legal separation (i.e. horizontally 5 nautical miles and vertically 1000 at or below Flight Level 290 and 2000 feet above). Encounters are assumed at larger separations both horizontally and vertically (i.e. horizontally 30 nautical miles and vertically 4000/5000 feet).

relative impact flight intent had on the conflict prediction accuracy is separated from the impact induced by weather forecast errors.

The other criteria considered in the analyses are the alert types. URET presents red, yellow, muted red, and muted yellow alerts. First, both muted red and yellow alerts are excluded from all the analyses. For this study and consistent with the URET CCLD Formal Accuracy Test, muted alerts are not considered valid alerts and will effectively end a notification (e.g. a red alert changing to a muted red alert effectively terminates the notification). In regards to the URET CCLD Formal Test, these rules are explained in detail in [LMATM, 2001] and applied in [LMATM, 2002]. The analyses AA and BA considered only the URET red alerts. The analyses AB and BB considered both the red and yellow alerts. URET red alerts are conflict predictions that have predicted minimum separations less than five nautical miles. Yellow alerts are conflict predictions with predicted minimum separations of less than 12 nautical miles. The red alerts more closely match what would be expected as a standard legal aircraft to aircraft conflict, while the yellow alerts are predicting conflicts at larger or less critical separations. Since a red alert predicts smaller separations, it is expected that false alerts for the red only analyses would be less compared to the red and yellow alert analyses. However, the missed alerts for the red only analyses would be expected to be higher for the same reason.

Since the analysis uses the same air traffic scenario for all the runs, the ground truth information will be the same in evaluation of the conflict predictions. The HCS messages are mainly used to determine the ground truth (i.e. actual position of the aircraft, intent of the aircraft, and any conflicts between aircraft), which are extracted from the scenario. Thus, the same conflicts are evaluated for all the runs and only the predictions are potentially different.

2.2.3.1 Missed, Valid, and False Alerts

When URET predicts that a future conflict will occur between two aircraft, it posts an alert to the air traffic controller's display. The alert remains posted until the conflict is past or is no longer predicted. Usually the controller will redirect one of the aircraft so that the conflict will not occur. URET automatically reads this change in flight path and deletes the alert. The alert may be updated (in time or space), while it is posted to the controller's display. The initial posting of the alert and its final deletion form a notification set which can be matched to an actual conflict.

A CP, like URET, is not perfect and does make mistakes in its conflict predictions. To quantify these errors, the conflict prediction accuracy metrics describe two fundamental events: a conflict and an alert. These events, which are not mutually exclusive, have four possible outcomes (see Table 2.2-6). The conflict accuracy metrics quantify the two fundamental error outcomes: missed alert and false alert. CPAT first defined these errors and rules to measure them in [Cale et al., 1998], but others have applied similar techniques in [Brudnicki, 1998] and [Bilimoria, 2001].

Table 2.2-6: URET Alert and Conflict Event Combinations

	CONFLICT OCCURS	CONFLICT DOES NOT OCCUR
ALERT	URET predicts conflict and it occurs (V -- valid alerts)	URET predicts conflict and it does not occur (F -- false alert)
NO ALERT	URET does not predict conflict and it occurs (M -- missed alert)	URET does not predict conflict and it does not occur (NC -- correct no-calls)
Totals	Total Number of Conflicts	Total Number of Non-Conflicts (Encounters that did not have conflicts)

For a real time system, it is important that an alert be given in sufficient time prior to the actual conflict so corrective action can be taken. In other words, an alert must be timely as well as accurate. Under normal conditions in the URET CCLD Formal Accuracy Test, the FAA's strategic conflict probe was required to have a five-minute lead-time or actual warning time. A notification set is evaluated as a valid alert when URET correctly predicts the conflict and when it is posted in a timely manner. If the notification set is not presented at all or correctly predicts the conflict but is not posted soon enough, it is called a missed alert. The lateness of the alert may be excused only if the conflict is considered a pop-up, which is defined in detail in [Paglione et al., 2002] and [LMATM, 1998]. Briefly, a pop-up conflict occurs if URET is not provided with the same five-minute time threshold of continuous HCS data or prediction for either of the associated flights. A notification set determined to be a missed alert due to lateness is also referred to as a late missed alert or strategic missed alert. A notification set presented late but excused is referred to as a late valid alert. A notification set that predicts a conflict when no conflict occurs is a false alert. However, an alert withdrawn before the predicted conflict start time is called a retracted false alert. A false alert is not matched to a conflict but an encounter and may be excused as well.

Besides pop-up conflicts that allow relaxation of the five-minute timeliness threshold for valid alerts, adherence is a technique to filter out conflict probe accuracy data with erroneous flight intent. Using a concept called adherence age, both missed and false alerts may be discarded. This can occur if the associated flights are lacking flight intent data at either the missed alert's conflict start time or false alert's predicted conflict start time. A more detailed description of adherence is provided in [Paglione et al., 2000].

2.2.3.1.1 Methodology of Measuring Missed and False Alerts

As first presented in [Paglione et al., 2002], an effective way to present the methodology of measuring the conflict prediction accuracy is to describe the specific process used to quantify these error events. The missed, valid, and false alerts, as defined in Table 2.2-6, are determined in two sub-processes. In Process A (see Figure 2.2-5), conflicts are evaluated in order of actual conflict start time and matched against eligible notification sets. To be eligible for matching to a specific conflict, a notification set must have a posting time prior to the start of the conflict and must have an end or delete time after the start of the actual conflict. Thus, the notification must precede the conflict and must be active at the start of the conflict. The result is a new listing of valid alerts, missed alerts, and discarded conflicts.

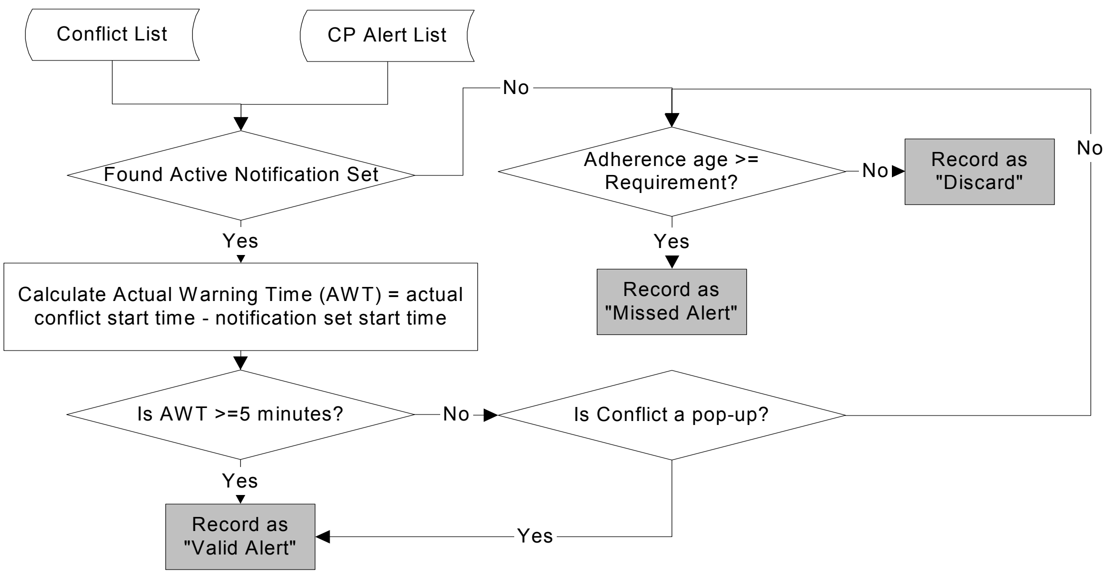


Figure 2.2-5: Process A - Valid and Missed Alert Processing

The discarded conflicts are either conflicts that have no eligible notification set or were notified late (i.e. a strategic missed alert) but can be discarded due to low adherence age. For Process A, only lack of adherence can excuse a missed alert, while adherence is not even checked if a valid alert is determined. In other words, if a CP, like URET, correctly predicts a conflict, the aircraft pair's flight intent is irrelevant. This allows a CP with superior heuristics that handle out of adherence situations more effectively to achieve higher accuracy results.

The remaining notification sets not matched as either valid alerts or discarded conflicts are potentially false alerts. In Process B (see Figure 2.2-6), the remaining notification sets are evaluated to determine which of them are truly false alerts and which can be discarded. Unlike the missed alerts, there are several reasons for discarding false alerts. The potential false alert is discarded if either aircraft does not have HCS track data present at the predicted conflict start time (PCST). With a lack of HCS track data, the false alert error is unverifiable and thus excused. In many of these cases, the discarded notification sets represent alerts predicted beyond the end of the traffic scenario.

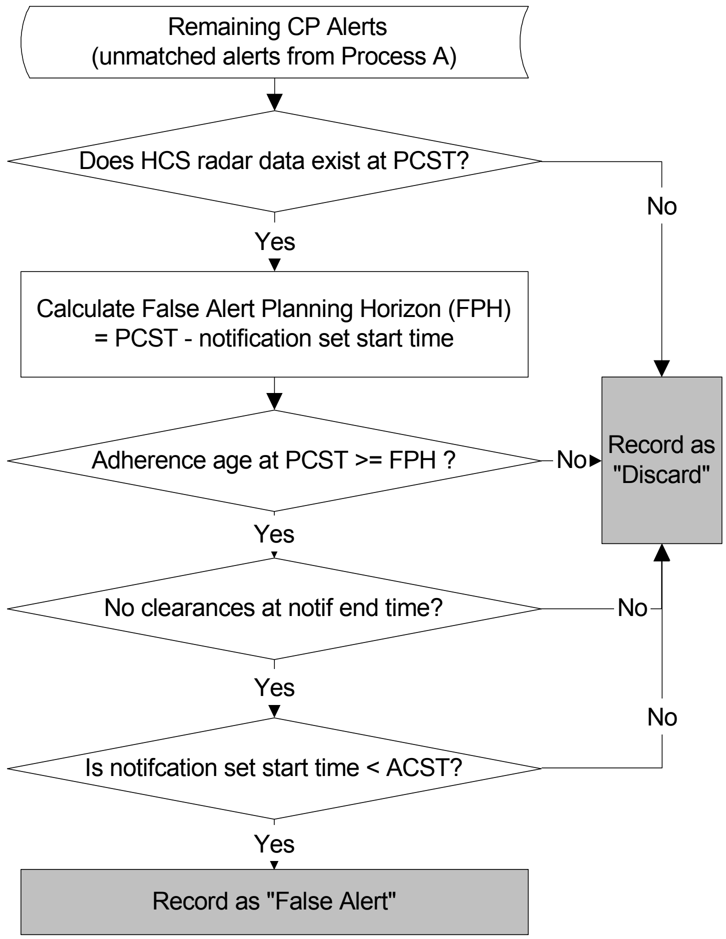


Figure 2.2-6: Process B - False Alert Processing

The potential false alert is discarded if either aircraft has a low adherence age at the predicted conflict start time. As discussed in the previous Section Definition of Adherence Age, this notification set may be discarded, when the involved aircraft have inadequate flight intent information available within the planning horizon of the prediction.

If the potential false alert is retracted due to an air traffic control clearance, the notification set is discarded. The potential false alert can also be discarded if the notification set was posted after the last actual conflict start time (ACST) between the associated aircraft. This can only happen if a conflict actually occurs between these aircraft and another alert is presented after it starts. When the URET is operating in the National Airspace System (NAS), once the actual conflict started, strategic alerts would have little value and other more tactical procedures would be utilized.

2.2.3.1.2 Reason Codes for Missed, False, and Valid Alerts

From Figure 2.2-5's Process A and Figure 2.2-6's Process B, there are four outcomes or alert types to the conflict prediction analysis. Initially, the alerts are matched with conflicts to produce valid, missed, or discarded alerts (Process A). The remaining alerts are evaluated as a false or again discarded (Process B). There are two subtypes for each of the valid, missed, and false alerts and several types of discarded alerts. These different subtypes are referred to as reason codes and are listed in the following Table 2.2-7.

Table 2.2-7: Reason Codes for Conflict Prediction Types[9]

REASON CODE	ALERT TYPE	REASON DESCRIPTION
STD_VA	Valid Alert	Standard Valid Alert
LATE_VA	Valid Alert	Late Valid Alert, Valid since conflict was determined a pop-up
NO_CALL_MA	Missed Alert	Missed Alert due to no call (no alert at all)
LATE_MA	Missed Alert	Late Missed Alert
NO_CALL_DISCARD	Discard Alert	Missed Alert no call discarded since out of adherence
LATE_DISCARD	Discard Alert	Late Discard since out of adherence
NO_TRK_FA_DISCARD	Discard Alert	No post processed track at predicted conflict start time so discard
NO_ADHER_FA_DISCARD	Discard Alert	Out of adherence at predicted conflict start time so discard
CLR_FA_DISCARD	Discard Alert	Retracted False Alert assigned by an ATC clearance so discard
CFL_FA_DISCARD	Discard Alert	False Alert notified beyond last conflict actual start time so discard
STD_FA	False Alert	Standard False Alert
RETRACT_FA	False Alert	Retracted False Alert, notification end time earlier than predicted conflict start time

2.2.3.2 Conflict Prediction Run Comparison

The sets of conflict predictions generated by the control run and a treatment run are evaluated separately. Then these evaluations are compared, which is only meaningful when both runs are provided the same input traffic scenario. The first column in Table 2.2-8 lists all combinations of intersection and union of the events from Table 2.2-6. This list is expanded in Table 2.2-9 to consider the discard events as well. Throughout this section, the first run compared will be referred to as Run A and the second as Run B.

[9] The discard rules were originally developed for the URET CCLD Formal Accuracy Test Program but will be applied to this analysis as well. They are defined in detail in [LMATM, 2001].

Table 2.2-8: Comparison of Two Run's Resulting Alert and Conflict Event Combinations

	CONFLICT OCCURS	CONFLICT DOES NOT OCCUR
ALERT by both Runs A and B	Both predicts conflict and it occurs (V_{A1}=V_{B1}-- valid alerts both)	Both predicts conflict and it does not occur (F_{A1}=F_{B1}-- false alert both)
ALERT by A and not B	A predicts conflict and it occurs (V_{A2} -- valid alerts by A only)	A predicts conflict and it does not occur (F_{A2} -- false alert by A only)
	B does not predict conflict and it occurs (M_{B2} -- missed alert by B only)	B does not predict conflict and it does not occur (NC_B -- correct no-calls by B only)
ALERT by B and not A	B predicts conflict and it occurs (V_{B2} -- valid alerts by B only)	B predicts conflict and it does not occur (F_{B2} -- false alert by B only)
	A does not predict conflict and it occurs (M_{A2} -- missed alert by A only)	A does not predict conflict and it does not occur (NC_A -- correct no-calls by A only)
NO ALERT by both Runs A and B	Both do not predict conflict and it occurs (M_{A1}= M_{B1} -- missed alert by both)	Both do not predict conflict and it does not occur (NC -- correct no-calls by both)
Total Number of Alerts for each/both	Total Number of Conflicts *(Same for both Runs!)*	Total Number of Non-Conflicts (Encounters that did not have conflicts; *Same for both Runs!*)

The situations described in Table 2.2-8 are thorough but not exhaustive. Table 2.2-8 does not include all the possible events when you consider the rules applied to determine the missed, false, and valid alerts. Specifically, for missed and valid alert combinations flight plan adherence could exclude certain missed alerts and thus conflicts under Run A and not under Run B. In other words, Run B could have successfully predicted the particular conflict resulting in a valid alert, while Run A did not. For Run A under this situation, the adherence rule allowed the missed alert to be discarded. Other discard rules are applied as well, particularly for false alerts. Therefore, Table 2.2-9 expands upon the situations listed in Table 2.2-8 even further to include all discard cases.

Table 2.2-9: Comparison of Two Run's Resulting Alert and Conflict Event Combinations With Discard Events

	CONFLICT OCCURS	CONFLICT DOES NOT OCCUR
ALERT by both Runs A and B	Both predicts conflict and it occurs (V_{A1}=V_{B1}-- valid alerts both)	Both predicts conflict and it does not occur (F_{A1}=F_{B1}-- false alert both)
ALERT by A and not B	A predicts conflict and it occurs (V_{A2} -- valid alerts by A only)	A predicts conflict and it does not occur (F_{A2} -- false alert by A only)
	B does not predict conflict and it occurs (M_{B2} -- missed alert by B only)	B does not predict conflict and it does not occur (NC_B -- correct no-calls/discards B only)
ALERT by A and B ALERT or non-ALERT is discarded	A predicts conflict and it occurs (V_{A3} -- valid alerts by A only)	A predicts conflict and it does not occur (** F_{A2} Continued **)
	B does not predict conflict correctly but is discarded ($Discard_B$ -- B discards only)	B does not predict conflict correctly but is discarded (** NC_B Continued **)
ALERT by B and not A	B predicts conflict and it occurs (V_{B2} -- valid alerts by B only)	B predicts conflict and it does not occur (F_{B2} -- false alert by B only)
	A does not predict conflict and it occurs (M_{A2} -- missed alert by A only)	A does not predict conflict and it does not occur (NC_A -- correct no-calls/discards A only)
ALERT by B and A ALERT or non-ALERT is discarded	B predicts conflict and it occurs (V_{B3} -- valid alerts by B only)	B predicts conflict and it does not occur (** F_{B2} Continued **)
	A does not predict conflict correctly but is discarded ($Discard_A$ -- A discards only)	A does not predict conflict correctly but is discarded (** NC_A Continued **)
NO ALERT by both Runs A and B	Both do not predict conflict and it occurs (M_{A1}= M_{B1} -- missed alert by both)	Both do not predict conflict and it does not occur (NC -- correct no-calls by both)
Total Number of Alerts for each/both	Total Number of Conflicts *(Same for both Runs!)*	Total Number of Non-Conflicts (Encounters that did not have conflicts; *Same for both Runs!*)

The verifiable conflicts for Run A or Run B are slightly different due to the potential for discarding. For discarding missed alerts, the only rule that could possibly apply is the flight plan adherence of the true conflict. The conflict would have to have an adherence age beyond a parameter time at the start of conflict (e.g. 13 minutes). Valid alerts predicting these conflicts with adherence age less than the threshold time would be still be valid, while missed alerts with the same adherence age can be discarded. This prevents penalizing a conflict probe from predicting alerts correctly even if the input intent of the flights is in error. Thus, the number of verifiable conflicts for a given run is the composite of the valid and missed alerts. Equation (1) is the total number of verifiable conflicts for Run A (C_A), while Equation (2) is the same for Run B (C_B). The term verifiable reflects that the conflicts associated with discarded missed alerts are excluded. Equation (3) lists the quantity of all the verifiable conflicts for both.

$$C_A = V_A + M_A = V_{A1} + V_{A2} + V_{A3} + M_{A1} + M_{A2} \qquad \textbf{(1)}$$

$$C_B = V_B + M_B = V_{B1} + V_{B2} + V_{B3} + M_{B1} + M_{B2} \qquad \textbf{(2)}$$

Now, adding Equation (1) and (2) and subtracting the common valid and missed alerts, equates to:

$$C_{ALL} = C_A + C_B - V_{B1} - V_{B2} - M_B \qquad \textbf{(3)}$$

Table 2.2-10 summarizes the event count variables in Table 2.2-9. Each column, except the first referring to the total conflict counts per run, represents variables that are equivalent. For example, V_{A1} is equal to V_{B1}.

Table 2.2-10: Summary of Event Count Variables

Conflicts	**Common Valid Alerts**	**Valid A and Missed B**	**Valid A and Discard B**[10]	**Common Missed Alerts**	**Missed A and Valid B**	**Discard A and Valid B**[11]
C_A	V_{A1}	V_{A2}	V_{A3}	M_{A1}	M_{A2}	$Discard_A$
C_B	V_{B1}	M_{B2}	$Discard_B$	M_{B1}	V_{B2}	V_{B3}

To compare the two runs, the difference in missed alert probability is of interest. Equation (4) is the missed alert probability for Run A, and Equation (5) is the same for Run B.

$$\text{Run A Probability of Missed Alert} = \frac{M_A}{C_A} \qquad \textbf{(4)}$$

$$\text{Run B Probability of Missed Alert} = \frac{M_B}{C_B} \qquad \textbf{(5)}$$

[10] V_{A3} are valid alerts only if flight plan adherence is used to discard conflicts in B.
[11] V_{B3} are valid alerts only if flight plan adherence is used to discard conflicts in A.

Taking the difference of Equation (4) and (5) equates to:

$$\text{Missed Alert Probability Difference} = \frac{M_A C_B - M_B C_A}{C_A C_B},$$

so by canceling and substituting terms becomes:

$$\text{Missed Alert Probability Difference} = \frac{M_A V_B - M_B V_A}{C_A C_B} \qquad \textbf{(6)}$$

Equation (6) is the general equation of the difference between runs in missed alert probability. If the adherence discard rule is not applied, there will be no discarded conflicts in both runs and Equation (6) is simplified to the following Equation (7).

$$\text{If no adherence discard rule, then } C_A = C_B = C \text{ and } V_{A3} = V_{B3} = 0.$$

$$\text{Missed Alert Probability Difference} = \frac{M_{A2} - V_{A2}}{C} = \frac{V_{B2} - M_{B2}}{C} \qquad \textbf{(7)}$$

Analogous to the missed alert probabilities, false alert probabilities can be examined also. The difference is the adherence rule does discard some false alerts, but there are several other rules which can allow the discard of false alerts. These are listed in Table 2.2-11. The conditional false alert probabilities are listed in Equations (8) and (9) for Runs A and B, respectively. The difference in false alert probabilities between runs is listed in Equation (10).

$$\text{Run A Probability of False Alert} = \frac{F_A}{A_A} \qquad \textbf{(8)}$$

$$\text{Run B Probability of Missed Alert} = \frac{F_B}{A_B} \qquad \textbf{(9)}$$

Where F_A and F_B are the total number of false alerts for Run A and Run B, respectively, and A_A and A_B are the total number of alerts for Run A and Run B, respectively.

Taking the difference of Equation (8) and (9) equates to:

$$\text{False Alert Probability Difference} = \frac{F_A A_B - F_B A_A}{A_A A_B},$$

so by substituting and canceling terms becomes:

$$\text{False Alert Probability Difference} = \frac{F_A V_B - F_B V_A}{A_A A_B} \qquad \textbf{(10)}$$

With A_A being the quantity of alerts for Run A and A_B being the quantity of alerts for Run B, it is also necessary to find the total quantity of alerts for both runs, which is analogous to the total number of conflicts expressed in Equation (3). Equation (11) expresses the number of Run A verifiable alerts and Equation (12) lists the number of Run B verifiable alerts (verifiable since the discarded false alerts are excluded). Equation (13) expresses the union of these two runs or the total number of alerts for both runs.

$$A_A = V_A + F_A = V_{A1} + V_{A2} + V_{A3} + F_{A1} + F_{A2} \qquad \textbf{(11)}$$

$$A_B = V_B + F_B = V_{B1} + V_{B2} + V_{B3} + F_{B1} + F_{B2} \qquad \textbf{(12)}$$

Now, adding Equation (11) and (12) and subtracting the common valid and false alerts, equates to:

$$A_{ALL} = A_A + A_B - V_{B1} - F_{B1} = A_A + A_B - V_{A1} - F_{A1} \qquad \textbf{(13)}$$

Besides the missed and false alert differences, there are other quantities of interest. These additional comparison probabilities are summarized in Table 2.2-11.

Table 2.2-11: Additional Comparison Probabilities

Equation	Description	Equation Number
$\frac{M_{A1}}{C_{All}}$ or $\frac{M_{B1}}{C_{All}}$	Common missed alert probability, that is the probability that both runs had missed the conflict	**(14)**
$\frac{V_{A1}}{C_{All}}$ or $\frac{V_{B1}}{C_{All}}$	Common valid alert probability, that is the probability that both runs had correctly predicted the conflict	**(15)**
$\frac{V_{A2}}{C_{All}}$ or $\frac{M_{B2}}{C_{All}}$	Probability that Run A correctly called the conflict while Run B missed the conflict	**(16)**
$\frac{V_{B2}}{C_{All}}$ or $\frac{M_{A2}}{C_{All}}$	Probability that Run B correctly called the conflict while Run A missed the conflict	**(17)**
$\frac{F_{A1}}{A_{All}}$ or $\frac{F_{B1}}{A_{All}}$	Common conditional false alert probability, that is the probability that both runs had a falsely predicted a conflict	**(18)**

2.2.3.2.1 Statistical Significance of Missed and False Alert Differences

The most critical quantities to determine a statistical difference between runs are the missed alert probability and the false alert probability. The difference between these values is quantified in Equations (6) and (10), respectively. One approach to determine if the difference is statistically significance is to utilize a binomial distribution and perform a hypothesis test concerning the difference between population proportions [Devore, 2000]. However, this technique assumes that the respective runs are independent. For this study, each run is not independent, since they are run with the same air traffic scenario and altering weather files.

An alternative technique is presented in [Kachigan, 1986], utilizing categorical data analysis techniques. For categorical data analysis, we examine the difference in frequencies not proportions. For this study, the frequencies directly relating the missed and false alert probabilities include the counts of these events. Paired counts that are mutually exclusive and exhaustive, which is required for this test, occur when the error event occurs in one run and the correct event occurs in the other.

For the missed alert analysis, the count of interest is the missed alert count in Run A when simultaneously getting a valid alert in Run B or vice versa for the opposite case. These include the counts V_{A2} or M_{B2} compared to the V_{B2} or M_{A2}. Therefore, the count of valid alerts in Run A and simultaneous missed alerts in Run B is statistically compared to the count of valid alerts in Run B and simultaneous missed alerts in Run A. These counts should be equally likely if the two runs are statistically equivalent. Calculating the ratio of the squared difference between the expected value of each run and the observed value can test this hypothesis. If the hypothesis is true, this ratio will follow a chi-squared distribution or χ^2 with one degree of freedom.

The test statistic is as follows:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad \textbf{(19)}$$

where

O_i is the observed frequency in category i

E_i is the expected frequency in category i

k is the total number of categories

For this study, k is always two, since only paired runs are compared. For example, the observed frequencies are the extracted V_{A2}/M_{B2} and V_{B2}/M_{A2} counts for the two runs. Since the null hypothesis assumes both events are equally likely, both expected frequencies are equal and calculated from the following equation:

$$E_i = \frac{\sum_{j=1}^{k} O_j}{k} \qquad \textbf{(20)}$$

The resulting test statistic in Equation (19) can be expressed as a probability or P-value[12] by assuming a chi-squared distribution with one degree of freedom. For example, let's say we observe a $V_{A2}/M_{B2} = 8$ and a $V_{B2}/M_{A2} = 22$. The expected frequency from Equation (20) is 15 for both values, and the resulting test statistic from Equation (19) is 6.53. Therefore for this example exercise, the P-value is 0.011. This expresses that the hypothesis that these runs have equivalent missed alerts is only about one percent likely and provides evidence to reject the null hypothesis. For this test in the study, a P-value, which is less than 0.10, is considered sufficient to reject the hypothesis.

[12] Refer to footnote number 7 for more detail on the P-value probability. For convenience, the P-value is the probability of the null hypothesis has occurred, so a small P-value (less than 0.10) would indicate the null hypothesis unlikely and should be rejected.

False alert probabilities can be analyzed in an analogous way. For the false alert counts, the observed frequency of F_{A2}/NC_B and F_{B2}/NC_A are compared.

2.2.3.2.2 Combinations of Conflict Prediction Run Comparisons

To determine the various combinations of comparative events as defined in Table 2.2-9, CPAT wrote a software tool to extract them from the conflict prediction results of a pair of conflict probe runs. The program produces a database table of entries with evaluation codes for each of these events. A listing of these combinations and their corresponding codes are listed in the following Table 2.2-12. These events form the basis of all the conflict prediction analysis described previously in the subsections of Section 2.2.3.2.

Table 2.2-12: Conflict Prediction Comparison Program Evaluation Codes

Event	Evaluation Code	Description
V_{A1} or V_{B1}	SAME_VA	Both runs have valid alerts for the same conflict
M_{A1} or M_{A1}	SAME_MA	Both runs have missed alerts for the same conflict
F_{A1} or F_{B1}	SAME_FA	Both runs have false alerts for the same encounter
V_{A2} or M_{B2}	VA_MA	Run A has a valid alert and Run B has a missed alert for the same conflict
M_{A2} or V_{B2}	MA_VA	Run A has a missed alert and Run B has a valid alert for the same conflict
V_{A3} or $Discard_B$	VA_DISCARD	Run A has a valid alert while Run B discards the conflict
$Discard_A$ or V_{B3}	DISCARD_VA	Run A discards the conflict while Run B has a valid alert
F_{A2} or NC_B	FA_NC	Run A has a false alert while Run B either has no prediction or discards the alert for the same encounter
NC_A or F_{B2}	NC_FA	Run A either has no prediction or discards the alert while Run B has a false alert for the same encounter

2.2.3.3 Valid Alert Attribute Comparison

As described in Section 2.2.2.2.2 and coded in Table 2.2-12 as SAME_VA, the common valid alerts are determined for each run comparison (e.g. Run 000 versus Run 200 and coded 000_200). This means both the control run and the treatment run have a common valid alert for the particular conflict in question. In this section further analysis is performed on two attributes of these alerts, which include the warning time and conflict predicted start time deviations. The warning time is the lead-time in which the valid alert is displayed before the actual conflict start time. It is calculated as the difference in actual conflict start time minus notification start time. For the conflict start time error, the predicted conflict start time of the valid alert is subtracted by the actual conflict start time. This analysis will compare both warning times and conflict start time errors of the common valid alerts to determine if the treatment has a greater error in these statistics than the control run.

2.2.3.3.1 Definition of Warning and Start Time Comparisons

Once again, the warning time for a given valid alert is calculated by the following equation:

$$WT_i = ACST_i - NS_i \qquad \textbf{(21)}$$

where WT_i is the warning time of valid alert i, NS_i is the notification start time of valid i, and $ACST_i$ is the actual conflict start time of the matched conflict of valid alert i.

Now, comparing the Run A's (e.g. control run, Run 000) valid alert warning time with the Run B's (e.g. treatment run, Run 200) common valid alert warning time would include taking the difference of Equation (21) for each, such as:

$$\Delta WT_i = \left(ACST_i - NS_i\right)_{RunA} - \left(ACST_i - NS_i\right)_{RunB} \qquad \textbf{(22)}$$

where $ACST_i$ is the same for Run A and Run B, so the Equation (22) reduces to the following:

$$\Delta WT_i = \left(NS_i\right)_{RunB} - \left(NS_i\right)_{RunA} \qquad \textbf{(23)}$$

Equation 23 illustrates that the difference in warning time for a common valid alert is just the difference in notification start time. A positive value for Equation (23) indicates that Run B was presented later than Run A, thus with less warning time. A negative value for Equation (23) indicates the opposite (that is Run B provided more warning time than Run A).

Analogous to warning time, the start time error is the difference between the actual start time and the predicted conflict start time. It is expressed in the following Equation (24).

$$ST_i = ACST_i - PCST_i \qquad \textbf{(24)}$$

where ST_i is the start time deviation of valid alert i, $PCST_i$ is the predicted conflict start time of valid i, and $ACST_i$ is the actual conflict start time of the matched conflict of valid alert i.

Now, in the same manner as the warning time was compared, the following Equation (25) expresses the difference between a common valid alert's predicted conflict start time. A positive value from Equation (25) indicates the start time of Run B was predicted to be later than Run A. A negative value to Equation (25) indicates that the Run B's predicted start time was earlier than Run A.

$$\Delta ST_i = \left(PCST_i\right)_{RunB} - \left(PCST_i\right)_{RunA} \qquad \textbf{(25)}$$

2.2.3.3.2 Statistical Tests for Warning and Start Time Comparisons

The differences in warning times as expressed in Equation (23) and in start time deviation in Equation (25) need to be statistically tested for significance. The hypothesis test is to determine if the difference is equal to zero (e.g. $H_o : \Delta WT_i = 0$), which means the warning time between the Run A and B was not different. An equivalent hypothesis test would be applied for start time deviation. The statistical technique chosen to test these hypotheses is the Wilcoxon signed-rank test. This is a nonparametric approach and is explained thoroughly in Section 2.2.2.2.3 Nonparametric Statistical Experiment, where the technique is described in context of testing the difference in trajectory prediction accuracy measurements. For redundancy, the description of this technique will not be repeated here.

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

3 Analysis and Results

3.1 Trajectory Accuracy Results

A statistical analysis was completed on the STR-LEV (referred to as Level), and on the STR-TRN (referred to as In-Transition) phase-of-flight combinations only. Additionally, only the level group had sufficient observations to do an analysis beyond the 300 second look-ahead time. A determination of statistical significance was not straightforward and required consideration of the various measures provided in the SAS Univariate generated reports. The primary means to determine statistical significance was the Wilcoxon signed-rank test, however the median, quantiles, and interquartile range (i.e. range between 75th and 25th quantiles) were used to validate the conclusion. Specifically the test results were verified based on practical significance as shown in the quantiles and interquartile range.

The following sections provide tables summarizing results of the pairwise tests for statistical significance by each weather forecast factor, phase-of-flight (PoF) group, and look-ahead time. General comments for each error type at the zero look-ahead time are also provided.

3.1.1 Wind Magnitude

As described in Section 2.1.2, a wind magnitude error was induced in the 100 and 200 runs by adding a 20 or 60 knot value to all locations in the RUC file's wind magnitude forecast. The 000-100 data was generated by subtracting the measured trajectory deviation error in the control run by the error incurred in Run 100 (20 knot induced error). Next, the measured trajectory deviation error of the control run is subtracted by the Run 200 error, which is referred to as the 000-200 analysis. Finally, the difference between runs 100-200 was computed by comparing Run 100 versus Run 200.

3.1.1.1 Statistical Results

Table 3.1-1 (level PoF group) and Table 3.1-2 (in-transition PoF) provide summary conclusions by trajectory error type, run sequence and look-ahead time. A No designation in the tables indicates a conclusion that adding a positive 20 or 60 knots to the control RUC file had no statistically significant effect on measured trajectory deviation. A Yes designation indicates a conclusion that the altered file did affect the trajectory accuracy. There was no analysis beyond the 300 second look-ahead time for the in-transition PoF group do to insufficient sample size.

Table 3.1-1: Wind Magnitude Comparison Results for Level Flight

Error Type	Run Codes	Look-Ahead Time 0	300	600	900	1200
Horizontal Error	000-100	Yes	Yes	Yes	Yes	Yes
	000-200	Yes	Yes	Yes	Yes	Yes
	100-200	Yes	Yes	Yes	Yes	Yes
Vertical Error	000-100	No	No	No	No	No
	000-200	No	No	No	No	No
	100-200	No	No	No	No	No
Longitudinal Error	000-100	Yes	Yes	Yes	Yes	Yes
	000-200	Yes	Yes	Yes	Yes	Yes
	100-200	Yes	Yes	Yes	Yes	Yes
Lateral Error	000-100	No	No	No	No	No
	000-200	No	No	No	No	No
	100-200	No	No	No	No	No

Table 3.1-2: Wind Magnitude Comparison Results for In-Transition Flight

Error Type	Run Codes	Look-Ahead Time 0	300
Horizontal Error	000-100	Yes	Yes
	000-200	Yes	Yes
	100-200	Yes	Yes
Vertical Error	000-100	No	No
	000-200	No	No
	100-200	No	No
Longitudinal Error	000-100	Yes	Yes
	000-200	Yes	Yes
	100-200	Yes	Yes
Lateral Error	000-100	No	No
	000-200	No	No
	100-200	No	No

3.1.1.2 General Observations

Horizontal error was determined to be statistically significant from zero for both PoF groups. The SAS univariate reports show the Student-t test to routinely support the null hypothesis of no difference but these results are consistently invalidated by the various tests for normality. The signed-rank and sign tests both support the alternative hypothesis. The P-value from the signed-rank test for the 000-100, in-transition PoF for zero look-ahead time run was 0.0734 which is borderline support of the null hypothesis.

Vertical error was determined to be not significantly different from zero. The quantiles showed extensive zero observations inside of large positive and negative outliers. Quantiles for the in-transition PoF show fewer zeros, but the scale of the observations about the median made the data essentially equal to zero. This phenomena known as practical significance was defined in Section 2.2.2.2.6 on Applied Statistical Approach.

Longitudinal error was determined to be significantly different from zero. The Student-t test supported the null but this result is invalidated by tests for normality. The signed-rank and sign tests support the alternative hypothesis. Significance was further verified by comparing the interquartile range and quantiles with those of the lateral error. Where the interquartile range was at least a half nautical mile for the longitudinal error the interquartile range equaled zero for lateral error. Since one of the orthogonal components of horizontal error must be significant and lateral error will be shown to be not significant then longitudinal error is confirmed to be significant different from zero.

Lateral error was not significantly different from zero. The Student-t test consistently supported the null hypothesis but is again invalidated by normality tests. The signed-rank and sign tests both support the null hypothesis. The quantiles showed at least eighty percent of the observations to equal zero. The zeros do invalidate the Wilcoxon methods but logically support the null hypothesis of no difference.

It was additionally observed that variability in the error data increased with wind magnitude as indicated by an increase in both range and interquartile range. Specifically, the interquartile range is larger for the 000-200 run than for the 000-100 run and the 100-200 run is in-between the two.

3.1.2 Wind Direction

As described in Section 2.1.2 and similar to the previous analysis on wind magnitude, the wind direction quantity added to the control run's RUC file at each grid point was a positive 45 or 90 degrees. The analysis sequence 000-010 represents the difference in measured trajectory error between the control or Run 000 and Run 010 where the RUC file was altered by adding 45 degrees to each grid point. The 000-020 sequence is the difference between the control and Run 020. For this Run 020, the RUC file is altered by adding 90 degrees to each wind direction grid point. The 010-020 analysis sequence is computed by comparing the two altered files.

3.1.2.1 Statistical Results

Table 3.1-3 (level PoF) and Table 3.1-4 (in-transition PoF) show test results for wind direction by error type, run sequence and look-ahead time. A 'No' designation in the table indicates a conclusion that adding either 45 or 90 degrees to the control run's wind direction had no statistically significance affect on the measured trajectory deviation. A 'Yes' designation indicates that the added component was statistically significant from zero. No analysis was done beyond the 300 second look-ahead time for the in-transition PoF due to small sample size.

Table 3.1-3: Wind Direction Comparison Results for Level Flight

Error Type	Run Codes	Look-Ahead Time 0	300	600	900	1200
Horizontal Error	000-010	No	Yes	Yes	Yes	Yes
	000-020	Yes	Yes	Yes	Yes	Yes
	010-020	Yes	Yes	Yes	Yes	Yes
Vertical Error	000-010	No	No	No	No	No
	000-020	No	No	No	No	No
	010-020	No	No	No	No	No
Longitudinal Error	000-010	No	Yes	Yes	Yes	Yes
	000-020	No	Yes	Yes	Yes	Yes
	010-020	Yes	Yes	Yes	Yes	Yes
Lateral Error	000-010	No	No	No	No	No
	000-020	No	No	No	No	No
	010-020	No	No	No	No	No

Table 3.1-4: Wind Direction Comparison Results for In-Transition Flight

Error Type	Run Codes	Look-Ahead Time 0	300
Horizontal Error	000-010	No	No
	000-020	Yes	Yes
	010-020	Yes	Yes
Vertical Error	000-010	No	No
	000-020	No	No
	010-020	No	No
Longitudinal Error	000-010	No	No
	000-020	Yes	Yes
	010-020	Yes	Yes
Lateral Error	000-010	No	No
	000-020	No	No
	010-020	No	No

3.1.2.2 General Observations

Horizontal error is not significantly different from zero for the 000-010 run comparison, level PoF, and zero look-ahead time run only. All other runs and look-ahead time sequences were significantly different from zero. The same determination was made for the 000-010, in-transition PoF for the zero look-ahead time. The comparison data for all runs were negatively skewed (i.e. has a heavier negative tail), which for a control minus treatment data set would indicate larger error deviations in the altered data set. The interquartile range (i.e. range between 75th and 25th quantiles) is larger for the level phase-of-flight. The range is about 0.35 nm versus 0.20 nm for level and in-transition phase, respectively.

Vertical error is determined to be not significantly different from zero based on the extensive number of zero observations present in the data. Essentially 90 percent of the observations are zero for the level PoF group which logically supports the null hypothesis. Vertical error for the in-transition PoF show similar results with fewer zero observations, but here a conclusion also supporting the null is based on practical significance.

Longitudinal error was determined to be significantly different from zero for all run sequences and look-ahead times except for the 000-010 (for both PoF groups) and the 000-020, in-transition groups at look-ahead time zero. The interquartile range is around one-half nautical mile for the zero look-ahead time and increase sequentially to almost six nautical miles at the 1200 second look-ahead time. All data sets are negatively skewed (treatment error larger than nominal) but become less so with look-ahead time.

Lateral error had mixed conclusions with the significance tests supporting the null for the 000-010 and then the alternative hypothesis for the 000-020 and 010-020, level PoF group. In contrast, test results for the in-transition group strongly supported the null for all run sequences. Ultimately a determination supporting the null hypothesis was based on the extensive number of zero observations in the data.

3.1.3 Air Temperature

As described in Section 2.1.2, the air temperature factor added to the nominal level was a positive 5 or 15 degrees. The run sequence 000-001 represents the difference in measured trajectory error between the control run (Run 000) and the RUC file altered by adding 5 degrees to the temperature prediction at each grid point (Run 010). The run sequence 000-002 represents the difference between the control and the RUC file altered by adding 15 degrees. Finally, the run sequence 001-002 represents the difference in trajectory error between the two altered RUC files.

3.1.3.1 Statistical Results

Table 3.1-5 (level PoF) and Table 3.1-6 (in-transition PoF) provide the statistical results for air temperature. A ‘No’ designation in the tables indicates a conclusion that the difference in trajectory deviation is due to adding either 5 or 15 degrees to the nominal temperature in Run 000 was not statistically different from zero. A ‘Yes’ designation indicates a conclusion that the added temperature component was statistically significant. No analysis was done beyond the 300 second look-ahead time for the in-transition PoF due to small sample size.

Table 3.1-5: Air Temperature Comparison Results for Level Flight

Error Type	Run Codes	Look-ahead Time				
		0	300	600	900	1200
Horizontal Error	000-001	No	No	No	No	No
	000-002	No	No	No	No	No
	001-002	No	No	No	No	No
Vertical Error	000-001	No	No	No	No	No
	000-002	No	No	No	No	No
	001-002	No	No	No	No	No
Longitudinal Error	000-001	No	No	No	No	No
	000-002	No	No	No	No	No
	001-002	No	No	No	No	No
Lateral Error	000-001	No	No	No	No	No
	000-002	No	No	No	No	No
	001-002	No	No	No	No	No

Table 3.1-6: Air Temperature Comparison Results for In-Transition Flight

Error Type	Run Codes	Look-ahead Time	
		0	300
Horizontal Error	000-001	No	No
	000-002	Yes	No
	001-002	No	No
Vertical Error	000-001	No	No
	000-002	No	No
	001-002	No	No
Longitudinal Error	000-001	No	No
	000-002	Yes	No
	001-002	No	No
Lateral Error	000-001	No	No
	000-002	No	No
	001-002	No	No

3.1.3.2 General Observations

Horizontal error was determined to be not statistically different from zero except for the 000-002, in-transition run sequence. The 000-002 run will be discussed following general observations on the remaining runs. The median observation for the other runs was consistently equal to zero and the interquartile range was either zero or essentially equal to zero. All tests supported the null hypothesis of no difference in horizontal error deviation between runs though the Student-t test was invalidated by the various normality tests and the signed-rank test was invalidated do to extensive zero observations. The extensive zero observations logically support the null hypothesis.

Horizontal error for the 000-002, in-transition, zero look-ahead time run was statistically significantly for the both Wilcoxon methods (p=0.0277 for the signed-rank test) but not for the Student-t test (p=0.5750). The Student-t test result is invalidated by normality tests, but the signed-rank test might be valid based on the interquartile range of 0.1018 nautical miles. All the level PoF runs showed interquartile ranges close to zero. The other in-transition runs had interquartile ranges of 0.0721 (000-200) and 0.0338 (000-001) nautical miles. The 000-002 result is further validated by locating an orthogonal component (lateral or longitudinal error) that is also significant.

Vertical error was determined to be not significantly different from zero with a possible exception again being the 000-002 run sequence. All tests support the null though the normality tests invalidate the Student-t and the presence of extensive zero observations invalidate the signed-rank test. The extensive zeros and the application of practical significance in the in-transition PoF lead to a conclusion of not significantly different from zero.

Vertical error in the 000-002 run was probably not significant based on practical significance. The Student-t test strongly supported the null and both Wilcox methods supported the alternative hypothesis. The interquartile range of 79.25 feet was the largest of all runs but still on a scale that would make it not significantly different from zero.

Longitudinal error was not significantly different from zero for all sequences but the 000-002, in-transition, zero look-ahead time run. The Student-t supports the null but is again invalidated by normality tests. The signed-rank test strongly supports the alternative hypothesis for the 000-002 run. There are extensive zero observations in the reported quantiles for the level runs. There are fewer zeros in the in-transition runs. The interquartile range for the 000-002, in-transition run is 0.2087 nautical miles which is probably large enough to validate the signed-rank test result supporting the alternative hypothesis. The significant test result for longitudinal error (000-002) indicates this to be the orthogonal component to the statistically significant horizontal error.

Lateral error was determined to be not significantly different from zero in all run sequences including the 000-002 run do to extensive zero observations present in the data.

3.1.4 Trajectory Stability Results

As defined in [Lindsay, 1997a], trajectory stability indicates how often trajectories are rebuilt after being determined to be out of conformance. This occurs when an aircraft's track position is outside the region of uncertainty (conformance region) centered at the trajectory centerline. In this study, the total number of trajectories for each run is listed in Table 3.1-7. From this table, the greatest number of trajectories was generated in Run 200 and was followed by Run 020. This is consistent with the treatment runs with the highest trajectory errors as determined in the previous Sections 3.1.1 to 3.1.3.

Next, an analysis was performed that compared each treatment run against the control run. For each flight in the experiment, the number of trajectories generated from the control run was subtracted by the number of trajectories generated in the treatment run. Point statistics were generated for these differences and a Wilcoxon signed-rank test was performed. The hypothesis was tested that the differences were equal to zero, and thus the treatment runs had the same number of trajectories per flight as the control run. The results are listed in Table 3.1-8. For all the wind treatment runs (i.e. 100, 200, 010, and 020), the test was conclusive that their trajectory counts were statistically different than the control. A negative average in this table indicates that the treatment run has more trajectories than the control run. As listed in Table 3.1-8, all the wind magnitude runs had negative averages. For example, the Run 200 had on average 0.97 (almost one) more trajectories built per flight than the control run. The variability was also larger for the wind treatment runs, as indicated by the standard deviation and root mean square of differences listed in the Table 3.1-8. Therefore, for the wind magnitude and wind direction treatment runs, URET built more trajectories in total and for each flight as compared to the control run.

Table 3.1-7: Total Trajectory Count Per Run

Run Code	Total Trajectory Count
000	5156
100	5268
200	5622
010	5271
020	5501
001	5136
002	5181

Table 3.1-8: Trajectory Count Run Comparison Statistics

Comparative Statistics	Comparison Runs: Control Run Minus Treatment Run					
	000-100	000-200	000-010	000-020	000-001	000-002
Average	-0.46	-0.97	-0.24	-0.72	0.04	-0.05
Standard Deviation	1.26	1.68	1.49	1.89	0.71	1.09
Root Mean Square (RMS)	1.28	1.94	1.50	2.02	0.72	1.09
Sign-Rank Test: P-value	0.00	0.00	0.00	0.00	0.10	0.34

3.1.5 Summary of Trajectory Accuracy Results

The following paragraphs summarize results of the tests of hypothesis for testing whether adding a component to the RUC factors wind magnitude, wind direction or air temperature affects the measured trajectory deviation. Again, the measured deviation is the difference in trajectory deviation between a control run and a treatment run where the RUC factors were altered in the corresponding weather file.

Adding 20 or 60 knots to the wind magnitude factor was determined to affect horizontal error and longitudinal error (one of the two orthogonal components to the horizontal error) for both PoF groups (level and in-transition) considered in the study. The results for lateral error showed extensive zero observations which in a control minus treatment experiment logically support a conclusion of no difference between runs. Vertical error for the level PoF group also showed extensive zeros which supports a conclusion of no difference. The in-transition PoF group showed fewer zero observations but the scale of the observations about the median was so small to be practically insignificant.

Adding 45 or 90 degrees to the wind direction factor was determined to almost uniformly affect the horizontal and longitudinal trajectory deviation error for all look-ahead times. The single exception was at the lower factor level (000-010 run) for the zero look-ahead time. This contrary result was observed for both phase-of-flight groups. Additionally there was a conclusion of not significantly different from zero for the 000-020, level PoF, zero look-ahead time group but no corresponding test result for horizontal error. Both vertical and lateral error were determined to be not significantly different from zero based on the extensive zero observations present on the data.

Adding 5 or 15 degrees to the air temperature factor was determined to have no statistically significant effect on trajectory deviation error with the one exception being the 000-002 run at look-ahead time zero for the in-transition PoF group. Test results for horizontal error, the orthogonal components lateral and longitudinal error, and for vertical error were consistently not significant for all other groups and were further validated by extensive zero observations in the data. Test results at the upper factor level for the in-transition PoF group (000-002 run) showed significance for horizontal and longitudinal error. Lateral and vertical error for this run was not significantly different from zero.

3.2 Conflict Prediction Results

With the methodology described in detail in Section 2.2.3, the conflict prediction analysis for aircraft to aircraft conflict notifications includes a series of steps. The first is the determination of the missed, false, and valid alerts for each run of the experiment. Second is the comparison of these evaluated alerts from each run of the experiment (e.g. Run 000 to Run 200). This analysis step provides the bulk of the results as a series of comparative counts, differences in probabilities, and tests for statistical significance. The final step provides statistics on the difference in specific attributes of common conflict predictions. In other words, both runs being compared have correctly called the same conflict being examined, but this analysis determines how different these predictions are. The two attributes studied include warning time and conflict prediction start time. Section 3.2.1 will include the individual counts of missed, valid, and false alert events. Next in Section 3.2.2, the comparative statistics will be presented, and finally the statistics comparing the warning and conflict predicted start time are presented in Section 3.2.3.

As previously discussed in Section 2.2.3 and presented here again for simply for convenience, the analysis was performed four times under different criteria. Thus, four analyses were performed for each URET run and include:

- Analysis AA - adherence included and red alerts only
- Analysis BA - adherence ignored and red alerts only
- Analysis AB - adherence included and both red and yellow alerts
- Analysis BB - adherence ignored and both red and yellow alerts

As a review, the analyses considering adherence are expected to have better accuracy results compared to the analyses ignoring this criterion. By performing the analyses with and without adherence, the analyst can infer the relative difference in impacts from the lack of flight intent compared to the effect from the induced weather forecast errors of this study. By performing multiple analyses contrasting the URET red alert conflict predictions compared to the both red and yellow alerts, the analyst can infer the impact of the study’s induced weather forecast errors under these different criteria.

3.2.1 Individual Run's Missed, Valid, and False Alert Counts

This section presents each individual run's conflict prediction accuracy results. It includes the basic missed, valid, and false alert counts, but also their associated probabilities and the lower level counts of various reason codes defined in detail in Section 2.2.3.1.2.

3.2.1.1 Analysis AA's Individual Run Conflict Prediction Results

The following Table 3.2-1 illustrates the individual conflict prediction accuracy under the Analysis AA which considers only red alerts and requires adherence. The control run had a missed alert probability at 0.22 and a false alert probability at 0.53. The missed alert probabilities for all runs did not change much. The false alert probability did range from 0.51 to 0.58.

Table 3.2-1: Conflict Prediction Counts and Error Probabilities for Analysis AA

Description	Run Code						
	000	100	200	010	020	001	002
Actual Conflicts	211	211	211	211	211	211	211
Actual Encounters	1715	1715	1715	1715	1715	1715	1715
Alert Records	8036	8204	8687	8228	8479	7954	7875
Notification Sets	755	793	807	783	818	746	742
Missed Alerts	40	39	37	45	40	42	38
False Alerts	162	185	194	180	204	157	157
Valid Alerts	146	147	148	141	147	145	150
Discards	460	473	479	475	478	456	444
Verifiable Conflicts	186	186	185	186	187	187	188
Verifiable Alerts	308	332	342	321	351	302	307
Missed Alert Prob, P(M|C)	0.215	0.210	0.200	0.242	0.214	0.225	0.202
False Alert Prob, P(F|A)	0.526	0.557	0.567	0.561	0.581	0.520	0.511
Valid Alert Prob, P(V|A)	0.785	0.790	0.800	0.758	0.786	0.775	0.798

Table 3.2-2 presents the individual run's event reason code counts. From these reason codes, the dominating increase in false alerts from the control Run 000 is the retracted false alerts. This occurs only for the wind induced errors in Runs 100, 200, 010, and 020.

Table 3.2-2: Conflict Prediction Reason Code Counts for Analysis AA

Reason Code	Run Code						
	000	100	200	010	020	001	002
STD_VA	26	25	27	25	24	27	23
LATE_VA	120	122	121	116	123	118	127
NO_CALL_MA	29	27	28	33	29	31	26
LATE_MA	11	12	9	12	11	11	12
NO_CALL_DISCARD	24	24	23	25	22	23	21
LATE_DISCARD	1	1	3	0	2	1	2
NO_TRK_FA_DISCARD	198	196	203	208	212	193	189
NO_ADHER_FA_DISCARD	135	149	143	138	141	136	132
CLR_FA_DISCARD	61	65	68	65	58	64	65
CFL_FA_DISCARD	41	38	39	39	43	39	35
STD_FA	55	59	60	60	61	51	52
RETRACT_FA	107	126	134	120	143	106	105

3.2.1.2 Analysis BA's Individual Run Conflict Prediction Results

The following Table 3.2-3 illustrates the individual conflict prediction accuracy under the Analysis BA which considers only red alerts and ignores adherence. The control run had a missed alert probability at 0.31 and a false alert probability at 0.65. The missed alert probabilities for all runs did range from 0.29 to 0.33. The false alert probability ranged from 0.63 to 0.67.

Note, the notification set and valid alert counts are equivalent to the Analysis AA, but the false and missed alerts are higher and the discards less. The only difference between Analysis AA and BA is the former discards missed and false alerts due to adherence while this run does not.

Table 3.2-3: Conflict Prediction Counts and Error Probabilities for Analysis BA

Description	Run Code						
	000	100	200	010	020	001	002
Actual Conflicts	211	211	211	211	211	211	211
Actual Encounters	1715	1715	1715	1715	1715	1715	1715
Alert Records	8036	8204	8687	8228	8479	7954	7875
Notification Sets	755	793	807	783	818	746	742
Missed Alerts	65	64	63	70	64	66	61
False Alerts	265	299	299	278	304	264	257
Valid Alerts	146	147	148	141	147	145	150
Discards	332	334	348	352	354	325	321
Verifiable Conflicts	211	211	211	211	211	211	211
Verifiable Alerts	411	446	447	451	419	407	409
Missed Alert Prob, P(M|C)	0.308	0.303	0.299	0.303	0.332	0.289	0.313
False Alert Prob, P(F|A)	0.645	0.670	0.669	0.674	0.663	0.631	0.645
Valid Alert Prob, P(V|A)	0.692	0.697	0.701	0.697	0.668	0.711	0.687

Table 3.2-4 presents the individual run's event reason code counts. From these reason codes, the dominating increase in false alerts from the control Run 000 is the retracted false alerts. This only occurs for the wind induced errors in Runs 100, 200, 010, and 020.

Table 3.2-4: Conflict Prediction Reason Code Counts for Analysis BA

Reason Code	Run Code 000	100	200	010	020	001	002
STD_VA	26	25	27	25	24	27	23
LATE_VA	120	122	121	116	123	118	127
NO_CALL_MA	53	51	51	58	51	54	47
LATE_MA	12	13	12	12	13	12	14
NO_CALL_DISCARD	0	0	0	0	0	0	0
LATE_DISCARD	0	0	0	0	0	0	0
NO_TRK_FA_DISCARD	198	196	203	208	212	193	189
NO_ADHER_FA_DISCARD	0	0	0	0	0	0	0
CLR_FA_DISCARD	92	99	105	102	97	93	94
CFL_FA_DISCARD	42	39	40	42	45	39	38
STD_FA	70	75	74	74	74	69	64
RETRACT_FA	195	224	225	204	230	195	193

3.2.1.3 Analysis AB's Individual Run Conflict Prediction Results

The following Table 3.2-5 illustrates the individual conflict prediction accuracy under the Analysis AB which considers red and yellow alerts and includes adherence. The control run had a missed alert probability at 0.09 and a false alert probability at 0.63. The missed alert probabilities for all runs did not change much. The false alert probability did range from 0.62 to 0.67. Note, the notification set count is significantly higher for each run compared to the Analysis AA and BA, since now both red and yellow alerts are included.

Table 3.2-5: Conflict Prediction Counts and Error Probabilities for Analysis AB

Description	Run Code 000	100	200	010	020	001	002
Actual Conflicts	211	211	211	211	211	211	211
Actual Encounters	1715	1715	1715	1715	1715	1715	1715
Alert Records	8036	8204	8687	8228	8479	7954	7875
Notification Sets	1209	1230	1295	1222	1259	1195	1178
Missed Alerts	17	18	21	20	19	19	16
False Alerts	295	308	333	308	312	284	286
Valid Alerts	174	171	168	170	174	173	176
Discards	759	769	812	762	789	755	731
Verifiable Conflicts	191	189	189	190	193	192	192
Verifiable Alerts	469	479	501	478	486	457	462
Missed Alert Prob, P(M|C)	0.089	0.095	0.111	0.105	0.098	0.099	0.083
False Alert Prob, P(F|A)	0.629	0.643	0.665	0.644	0.642	0.621	0.619
Valid Alert Prob, P(V|A)	0.911	0.905	0.889	0.895	0.902	0.901	0.917

Table 3.2-6 presents the individual run's event reason code counts. From these reason codes, the noticeable increase in false alerts from the control Run 000 is the retracted false alerts. Once again, this only occurs for the wind induced errors in Runs 100, 200, 010, and 020.

Table 3.2-6: Conflict Prediction Reason Code Counts for Analysis AB

Reason Code	Run Code						
	000	100	200	010	020	001	002
STD_VA	41	41	42	41	41	42	39
LATE_VA	133	130	126	129	133	131	137
NO_CALL_MA	17	15	18	18	18	18	14
LATE_MA	0	3	3	2	1	1	2
NO_CALL_DISCARD	19	21	21	20	17	18	18
LATE_DISCARD	1	1	1	1	1	1	1
NO_TRK_FA_DISCARD	355	348	360	355	372	361	354
NO_ADHER_FA_DISCARD	226	244	266	237	246	227	214
CLR_FA_DISCARD	112	110	119	108	107	108	107
CFL_FA_DISCARD	46	45	45	41	46	40	37
STD_FA	161	158	164	163	157	150	146
RETRACT_FA	134	150	169	145	155	134	140

3.2.1.4 Analysis BB's Individual Run Conflict Prediction Results

The following Table 3.2-7 illustrates the individual conflict prediction accuracy under the Analysis BB which considers red and yellow alerts and ignores adherence. The control run had a missed alert probability at 0.18 and a false alert probability at 0.72. The missed alert probabilities for all runs did range from 0.17 to 0.20. The false alert probability did range from 0.71 to 0.76.

Note, the notification set and valid alert counts are equivalent to the Analysis AB, but the false and missed alerts are higher and the discards less. The only difference between Analysis AB and BB is the later discards missed and false alerts due to adherence while this run does not.

Table 3.2-7: Conflict Prediction Counts and Error Probabilities for Analysis BB

Description	Run Code						
	000	100	200	010	020	001	002
Actual Conflicts	211	211	211	211	211	211	211
Actual Encounters	1715	1715	1715	1715	1715	1715	1715
Alert Records	8036	8204	8687	8228	8479	7954	7875
Notification Sets	1209	1230	1295	1222	1259	1195	1178
Missed Alerts	37	40	43	41	37	38	35
False Alerts	449	485	522	475	483	440	429
Valid Alerts	174	171	168	170	174	173	176
Discards	585	570	601	574	600	580	569
Verifiable Conflicts	211	211	211	211	211	211	211
Verifiable Alerts	623	656	690	645	657	613	605
Missed Alert Prob, P(M|C)	0.175	0.190	0.204	0.194	0.175	0.180	0.166
False Alert Prob, P(F|A)	0.721	0.739	0.757	0.736	0.735	0.718	0.709
Valid Alert Prob, P(V|A)	0.825	0.810	0.796	0.806	0.825	0.820	0.834

Table 3.2-8 presents the individual run's event reason code counts. From these reason codes, the noticeable increase in false alerts from the control Run 000 is the retracted false alerts. Once again, this only occurs for the wind induced errors in Runs 100, 200, 010, and 020.

Table 3.2-8: Conflict Prediction Reason Code Counts for Analysis BB

Reason Code	Run Code						
	000	100	200	010	020	001	002
STD_VA	41	41	42	41	41	42	39
LATE_VA	133	130	126	129	133	131	127
NO_CALL_MA	36	36	39	38	35	36	32
LATE_MA	1	4	4	3	2	2	3
NO_CALL_DISCARD	0	0	0	0	0	0	0
LATE_DISCARD	0	0	0	0	0	0	0
NO_TRK_FA_DISCARD	355	348	360	355	372	361	354
NO_ADHER_FA_DISCARD	0	0	0	0	0	0	0
CLR_FA_DISCARD	183	176	194	175	179	179	175
CFL_FA_DISCARD	47	46	47	44	49	40	40
STD_FA	203	204	209	204	198	198	187
RETRACT_FA	246	281	313	271	285	242	242

3.2.2 Results on Conflict Prediction Run Comparison

This section compares runs as defined in Section 2.2.3.2. Each treatment run's conflict prediction events are compared to the control run (Run 000) forming six comparisons in all. In Section 2.2.3.2 Equations (6) and (10) are defined. These equations express the overall effect of the induced weather forecast errors to the missed and false alert probabilities. The Equations (14) through (18) are also calculated and present lower level differences between the runs. Next, a categorical statistical analysis is performed on the missed and false alert events as defined in Section 2.2.3.2.1. This determines if the missed and false alert counts between runs are statistically equivalent or indeed imply they are impacted by the treatment run's weather induced error. Furthermore, histograms are presented to illustrate the percentage increase or decrease in missed and false alert counts. Like the previous Section 3.2.1, the following subsections report the results for each of the four analyses and once again labeled as AA, AB, BA, and BB.

3.2.2.1 Analysis AA's Run Comparison Conflict Prediction Results

The following Table 3.2-9 illustrates the results of a conflict prediction accuracy comparison under the Analysis AA which considers red alerts only and requires adherence. The evaluations codes presented in this table are defined in the methodology Section 2.2.3.2.2. The difference in missed alert probabilities range from -0.03 to 0.02 without any specific pattern between runs. The false alert probability difference between the control and treatment runs ranges from -0.06 to 0.02. These false alert probabilities are higher for the wind treatment runs 100, 200, 010, and 020 only.

As defined in Section 2.2.3.2.1, the analysis answers two fundamental questions. First, the study determines whether the counts of VA_MA (i.e. number of valid alerts in the control run that are missed in the treatment run) is statistically equivalent to the MA_VA (i.e. number of missed alerts in the control run that are valid in the treatment run). This test provides evidence to state the weather forecast error did or did not induce a statistically equivalent missed alert error.

Similarly, the NC_FA and FA_NC counts are compared to determine if the false alert events were significantly impacted as well. Table 3.2-10 provides statistical evidence that the false alert probabilities are higher for the wind treatment runs 100, 200, 010, and 020 and not higher for the air temperature runs 001 and 002. Table 3.2-11 provides statistical evidence that the missed alert events probabilities are equivalent between the control and treatment runs. In other words, for Analysis AA only the wind treatment runs have a statistically significant effect on the false alert probability. However, the actual effect was not much in terms of false alert probability. The highest was evaluated for Run 020 where the false alert probability was about 0.06 higher than the control run. The percentage difference of false alert counts between the control and treatment runs is illustrated in Figure 3.2-1. It is calculated by taking the difference between the count of control run false alerts and the treatment false alert count and dividing this by the count of control run false alerts. A negative value would indicate that the treatment had more false alerts than the control run. From Figure 3.2-1, the retracted false alerts are responsible for the bulk of the differences with Run 020 having as much as 34 percent more false alerts than the control run.

Table 3.2-9: Conflict Prediction Run Comparison Statistics for Analysis AA

Comparative Counts							
Event	Evaluation Code	000-100	000-200		000-020	000-001	000-002
V_{A1} or V_{B1}	SAME_VA	135	134	134	135	141	142
M_{A1} or M_{B1}	SAME_MA	33	30	37	34	38	36
F_{A1} or F_{B1}	SAME_FA	124	110	121	112	130	112
V_{A2} or M_{B2}	VA_MA	6	7	8	6	4	2
M_{A2} or V_{B2}	MA_VA	7	10	3	6	2	4
V_{A3} or $Discard_B$	VA_DISCARD	5	5	4	5	1	2
$Discard_A$ or V_{B3}	DISCARD_VA	5	4	4	6	2	4
F_{A2} or NC_B	FA_NC	38	52	41	50	32	50
NC_A or F_{B2}	NC_FA	61	84	59	92	27	45
Comparative Statistics		000-100	000-200	000-010	000-020	000-001	000-002
$P(M_A\|C_A)$		0.215	0.215	0.215	0.215	0.215	0.215
C_A		186	186	186	186	186	186
$P(M_B\|C_B)$		0.210	0.200	0.242	0.214	0.225	0.202
C_B		186	185	186	187	187	188
$P(M_A\|C_A) - P(M_B\|C_B)$ [Eq. 6]		0.005	0.015	-0.027	0.001	-0.010	0.013
$P(F_A\|A_A)$		0.526	0.526	0.526	0.526	0.526	0.526
A_A		308	308	308	308	308	308
$P(F_B\|A_B)$		0.557	0.567	0.561	0.581	0.520	0.511
A_B		332	342	321	351	302	307
$P(F_A\|A_A) - P(F_B\|A_B)$ [Eq. 10]		-0.031	-0.041	-0.035	-0.055	0.006	0.015
C_{ALL} [Eq. 3]		191	190	190	192	188	190
A_{ALL} [Eq. 13]		381	406	374	412	339	361
$(M_{A1}$ or $M_{B1})/C_{ALL}$ [Eq. 14]		0.173	0.158	0.195	0.177	0.202	0.189
$(V_{A1}$ or $V_{B1})/C_{ALL}$ [Eq. 15]		0.707	0.705	0.705	0.703	0.750	0.747
$(V_{A2}$ or $M_{B2})/C_{ALL}$ [Eq. 16]		0.031	0.037	0.042	0.031	0.021	0.011
$(V_{B2}$ or $M_{A2})/C_{ALL}$ [Eq. 17]		0.037	0.053	0.016	0.031	0.011	0.021
$(F_{A1}$ or $F_{B1})/A_{ALL}$ [Eq. 18]		0.325	0.271	0.324	0.272	0.383	0.310

Table 3.2-10: Analysis AA False Alert Event Statistical Test

Statistics	Comparison Runs: Run A Versus Run B 000-100	000-200	000-010	000-020	000-001	000-002
FA_NC	38	52	41	50	32	50
NC_FA	61	84	59	92	27	45
Total	99	136	100	142	59	95
Expected	49.5	68	50	71	29.5	47.5
X^2	5.343	7.529	3.240	12.423	0.424	0.263
P-value	0.021	0.006	0.072	0.000	0.515	0.608

Table 3.2-11: Analysis AA Missed Alert Event Statistical Test

Statistics	Comparison Runs: Run A Versus Run B 000-100	000-200	000-010	000-020	000-001	000-002
VA_MA	6	7	8	6	4	2
MA_VA	7	10	3	6	2	4
Total	13	17	11	12	6	6
Expected	6.5	8.5	5.5	6	3	3
X^2	0.077	0.529	2.273	0.000	0.667	0.667
P-value	0.782	0.467	0.132	1.000	0.414	0.414

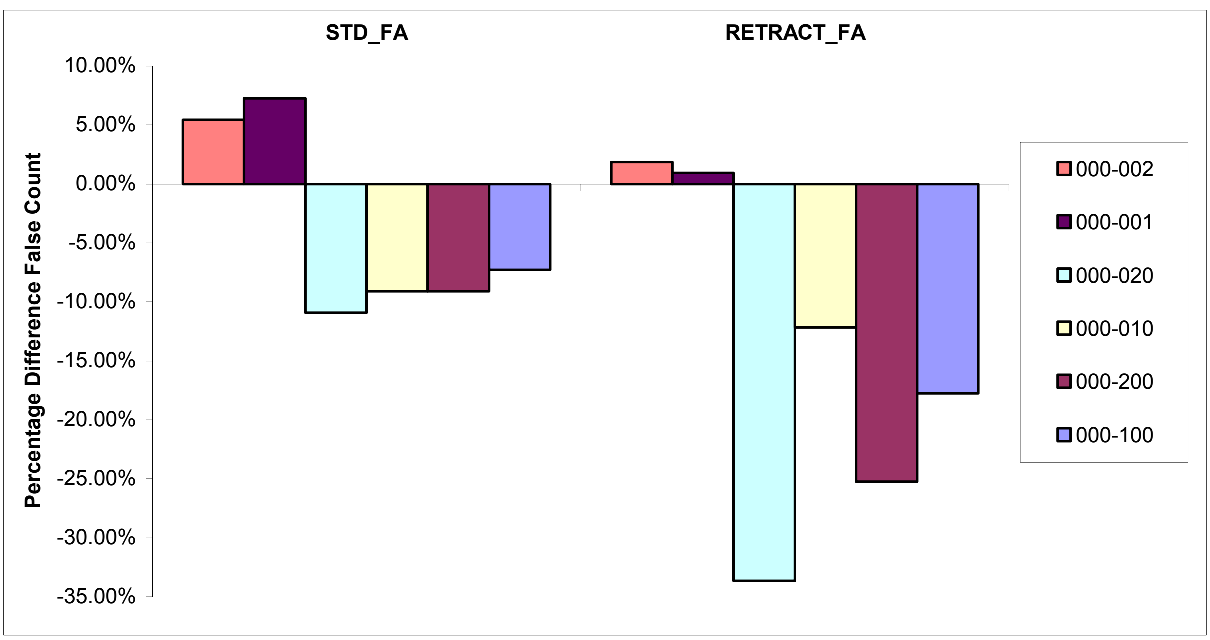


Figure 3.2-1: Analysis AA Run Percentage Difference of Standard and Retracted False Counts

3.2.2.2 Analysis BA's Run Comparison Conflict Prediction Results

The following Table 3.2-12 illustrates the results of a conflict prediction accuracy comparison under the Analysis BA which considers red alerts only and ignores adherence. Once again, the evaluations codes presented in this table are defined in the methodology Section 2.2.3.2.2. The difference in missed alert probabilities range from -0.02 to 0.02 without any specific pattern between runs. The false alert probability difference between the control and treatment runs ranges from -0.03 to 0.01. These false alert probabilities are consistently higher for the wind treatment runs 100, 200, 010, and 020 only.

Table 3.2-12: Conflict Prediction Run Comparison Statistics for Analysis BA

Comparative Counts		Comparison Runs: Run A Versus Run B					
Event	Evaluation Code	000-100	000-200	000-010	000-020	000-001	000-002
V_{A1} or V_{B1}	SAME_VA	135	134	134	135	141	142
M_{A1} or M_{B1}	SAME_MA	53	51	58	53	61	57
F_{A1} or F_{B1}	SAME_FA	210	178	196	180	222	193
V_{A2} or M_{B2}	VA_MA	11	12	12	11	5	4
M_{A2} or V_{B2}	MA_VA	12	14	7	12	4	8
V_{A3} or $Discard_B$	VA_DISCARD	0	0	0	0	0	0
$Discard_A$ or V_{B3}	DISCARD_VA	0	0	0	0	0	0
F_{A2} or NC_B	FA_NC	55	87	69	85	43	72
NC_A or F_{B2}	NC_FA	89	121	82	124	42	64
Comparative Statistics		000-100	000-200	000-010	000-020	000-001	000-002
$P(M_A\|C_A)$		0.308	0.308	0.308	0.308	0.308	0.308
C_A		211	211	211	211	211	211
$P(M_B\|C_B)$		0.303	0.299	0.332	0.303	0.313	0.289
C_B		211	211	211	211	211	211
$P(M_A\|C_A) - P(M_B\|C_B)$ [Eq. 6]		0.005	0.009	-0.024	0.005	-0.005	0.019
$P(F_A\|A_A)$		0.645	0.645	0.645	0.645	0.645	0.645
A_A		411	411	411	411	411	411
$P(F_B\|A_B)$		0.670	0.669	0.663	0.674	0.645	0.631
A_B		446	447	419	451	409	407
$P(F_A\|A_A) - P(F_B\|A_B)$ [Eq. 10]		-0.026	-0.024	-0.019	-0.029	-0.001	0.013
C_{ALL} [Eq. 3]		211	211	211	211	211	211
A_{ALL} [Eq. 13]		512	546	500	547	457	483
$(M_{A1}$ or $M_{B1})/ C_{ALL}$ [Eq. 14]		0.251	0.242	0.275	0.251	0.289	0.270
$(V_{A1}$ or $V_{B1})/ C_{ALL}$ [Eq. 15]		0.640	0.635	0.635	0.640	0.668	0.673
$(V_{A2}$ or $M_{B2})/ C_{ALL}$ [Eq. 16]		0.052	0.057	0.057	0.052	0.024	0.019
$(V_{B2}$ or $M_{A2})/ C_{ALL}$ [Eq. 17]		0.057	0.066	0.033	0.057	0.019	0.038
$(F_{A1}$ or $F_{B1})/ A_{ALL}$ [Eq. 18]		0.410	0.326	0.392	0.329	0.486	0.400

As defined in Section 2.2.3.2.1 and discussed in the previous Section 3.2.2.1, Table 3.2-13 provides statistical evidence that the false alert probabilities are higher for most of the wind treatment runs, namely 100, 200, and 020, and not higher for the air temperature runs 001 and

002. Table 3.2-14 provides statistical evidence that the missed alert events probabilities are equivalent between the control and treatment runs. In other words, for Analysis BA, only the wind treatment runs have a statistically significant effect on the false alert probability, but the actual effect was not very high in false alert probability with the largest evaluated for Run 020 (i.e. about 0.03 higher than the control run). Expressed as a percentage of the control run's false alert count, the difference in the false alert counts is illustrated in Figure 3.2-2. Thus, both standard and retracted false alerts are larger for the wind treatment runs, but the retracted false alerts seem to dominate.

Table 3.2-13: Analysis BA False Alert Event Statistical Test

	Comparison Runs: Run A Versus Run B					
Statistics	000-100	000-200	000-010	000-020	000-001	000-002
FA_NC	55	87	69	85	43	72
NC_FA	89	121	82	124	42	64
Total	144	208	151	209	85	136
Expected	72	104	75.5	104.5	42.5	68
X^2	8.028	5.558	1.119	7.278	0.012	0.471
P-value	0.005	0.018	0.290	0.007	0.914	0.493

Table 3.2-14: Analysis BA Missed Alert Event Statistical Test

	Comparison Runs: Run A Versus Run B					
Statistics	000-100	000-200	000-010	000-020	000-001	000-002
VA_MA	11	12	12	11	5	4
MA_VA	12	14	7	12	4	8
Total	23	26	19	23	9	12
Expected	11.5	13	9.5	11.5	4.5	6
X^2	0.043	0.154	1.316	0.043	0.111	1.333
P-value	0.835	0.695	0.251	0.835	0.739	0.248

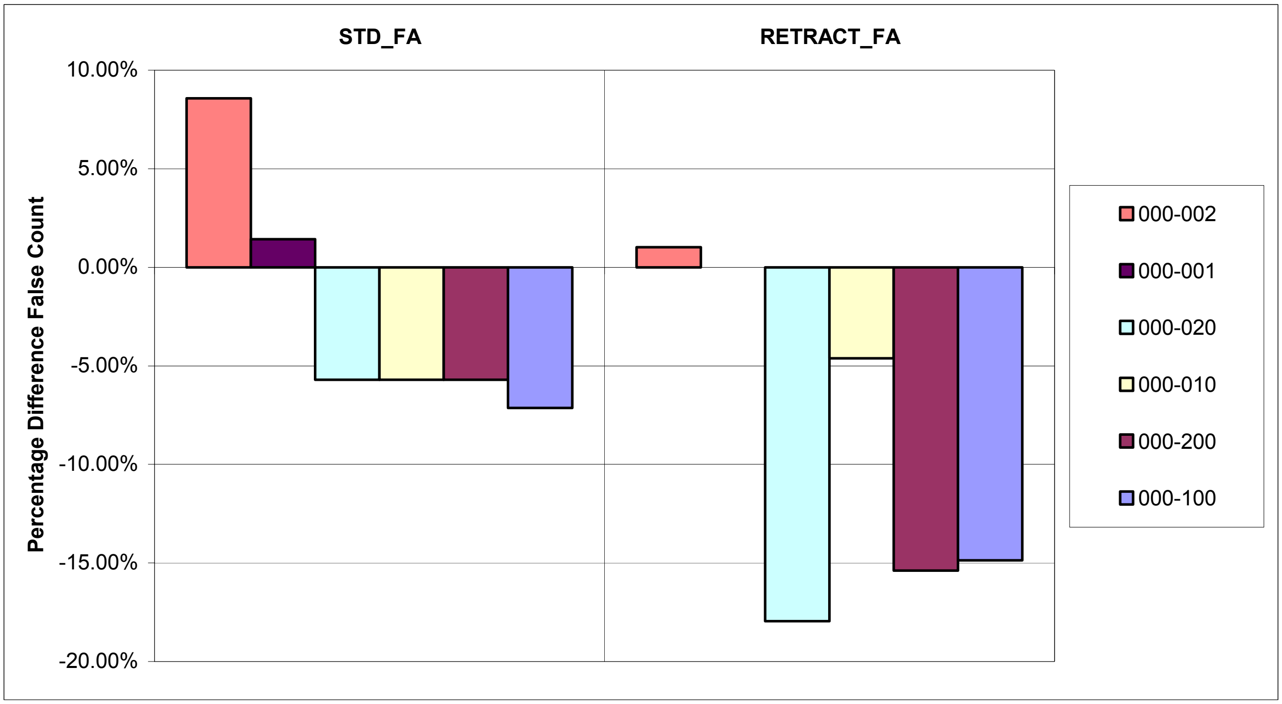


Figure 3.2-2: Analysis BA Run Percentage Difference of Standard and Retracted False Counts

3.2.2.3 Analysis AB's Run Comparison Conflict Prediction Results

The following Table 3.2-15 illustrates the results of a conflict prediction accuracy comparison under the Analysis AB which considers both red and yellow alerts and requires adherence. Once again, the evaluations codes presented in this table are defined in the methodology Section 2.2.3.2.2. The difference in missed alert probabilities range from -0.02 to 0.01 without any specific pattern between runs. The in false alert probability difference between the control and treatment runs ranges from -0.04 to 0.01. These false alert probabilities are higher for the wind treatment runs 100, 200, 010, and 020 only.

Table 3.2-15: Conflict Prediction Run Comparison Statistics for Analysis AB

Comparative Counts		Comparison Runs: Run A Versus Run B					
Event	Evaluation Code	000-100	000-200	000-010	000-020	000-001	000-002
V_{A1} or V_{B1}	SAME_VA	165	162	166	167	171	171
M_{A1} or M_{B1}	SAME_MA	14	13	15	14	16	14
F_{A1} or F_{B1}	SAME_FA	237	223	238	222	245	229
V_{A2} or M_{B2}	VA_MA	4	8	5	5	3	2
M_{A2} or V_{B2}	MA_VA	3	4	2	3	1	3
V_{A3} or $Discard_B$	VA_DISCARD	5	4	3	2	0	1
$Discard_A$ or V_{B3}	DISCARD_VA	3	2	2	4	1	2
F_{A2} or NC_B	FA_NC	58	72	57	73	50	66
NC_A or F_{B2}	NC_FA	71	110	70	90	39	57
Comparative Statistics		000-100	000-200	000-010	000-020	000-001	000-002
$P(M_A\|C_A)$		0.089	0.089	0.089	0.089	0.089	0.089
C_A		191	191	191	191	191	191
$P(M_B\|C_B)$		0.095	0.111	0.105	0.098	0.099	0.083
C_B		189	189	190	193	192	192
$P(M_A\|C_A) - P(M_B\|C_B)$ [Eq. 6]		-0.006	-0.022	-0.016	-0.009	-0.010	0.006
$P(F_A\|A_A)$		0.629	0.629	0.629	0.629	0.629	0.629
A_A		469	469	469	469	469	469
$P(F_B\|A_B)$		0.643	0.665	0.644	0.642	0.621	0.619
A_B		479	501	478	486	457	462
$P(F_A\|A_A) - P(F_B\|A_B)$ [Eq. 10]		-0.014	-0.036	-0.015	-0.013	0.008	0.010
C_{ALL} [Eq. 3]		194	193	193	195	192	193
A_{ALL} [Eq. 13]		546	585	543	566	510	531
$(M_{A1}$ or $M_{B1})/ C_{ALL}$ [Eq. 14]		0.072	0.067	0.078	0.072	0.083	0.073
$(V_{A1}$ or $V_{B1})/ C_{ALL}$ [Eq. 15]		0.851	0.839	0.860	0.856	0.891	0.886
$(V_{A2}$ or $M_{B2})/ C_{ALL}$ [Eq. 16]		0.021	0.041	0.026	0.026	0.016	0.010
$(V_{B2}$ or $M_{A2})/ C_{ALL}$ [Eq. 17]		0.015	0.021	0.010	0.015	0.005	0.016
$(F_{A1}$ or $F_{B1})/ A_{ALL}$ [Eq. 18]		0.434	0.381	0.438	0.392	0.480	0.431

As defined in Section 2.2.3.2.1 and discussed in the previous Section 3.2.2.1, the statistical hypothesis tests applied are presented in the Table 3.2-16 and Table 3.2-17. Table 3.2-16 provides statistical evidence that the false alert probabilities are higher for only one wind treatment run, Run 200, and not higher for the remaining wind treatment runs and air temperature runs. Table 3.2-17 provides statistical evidence that the missed alert events probabilities are equivalent between the control run and all the treatment runs. For Analysis AB only the wind treatment Run 200 has a statistically significant effect on the false alert probability. However, the actual effect was not much in terms of false alert probability with an increase of 0.04 from the control run. Expressed as a percentage of the control run's false alert count, the difference in the false alert counts is illustrated in Figure 3.2-3, which shows the retracted false alerts responsible for the increase.

Table 3.2-16: Analysis AB False Alert Event Statistical Test

	Comparison Runs: Run A Versus Run B					
Statistics	000-100	000-200	000-010	000-020	000-001	000-002
FA_NC	58	72	57	73	50	66
NC_FA	71	110	70	90	39	57
Total	129	182	127	163	89	123
Expected	64.5	91	63.5	81.5	44.5	61.5
X^2	1.310	7.934	1.331	1.773	1.360	0.659
P-value	0.252	0.005	0.249	0.183	0.244	0.417

Table 3.2-17: Analysis AB Missed Alert Event Statistical Test

	Comparison Runs: Run A Versus Run B					
Statistics	000-100	000-200	000-010	000-020		000-002
VA_MA	4	8	5	5	3	2
MA_VA	3	4	2	3	1	3
Total	7	12	7	8	4	5
Expected	3.5	6	3.5	4	2	2.5
X^2	0.143	1.333	1.286	0.500	1.000	0.200
P-value	0.705	0.248	0.257	0.480	0.317	0.655

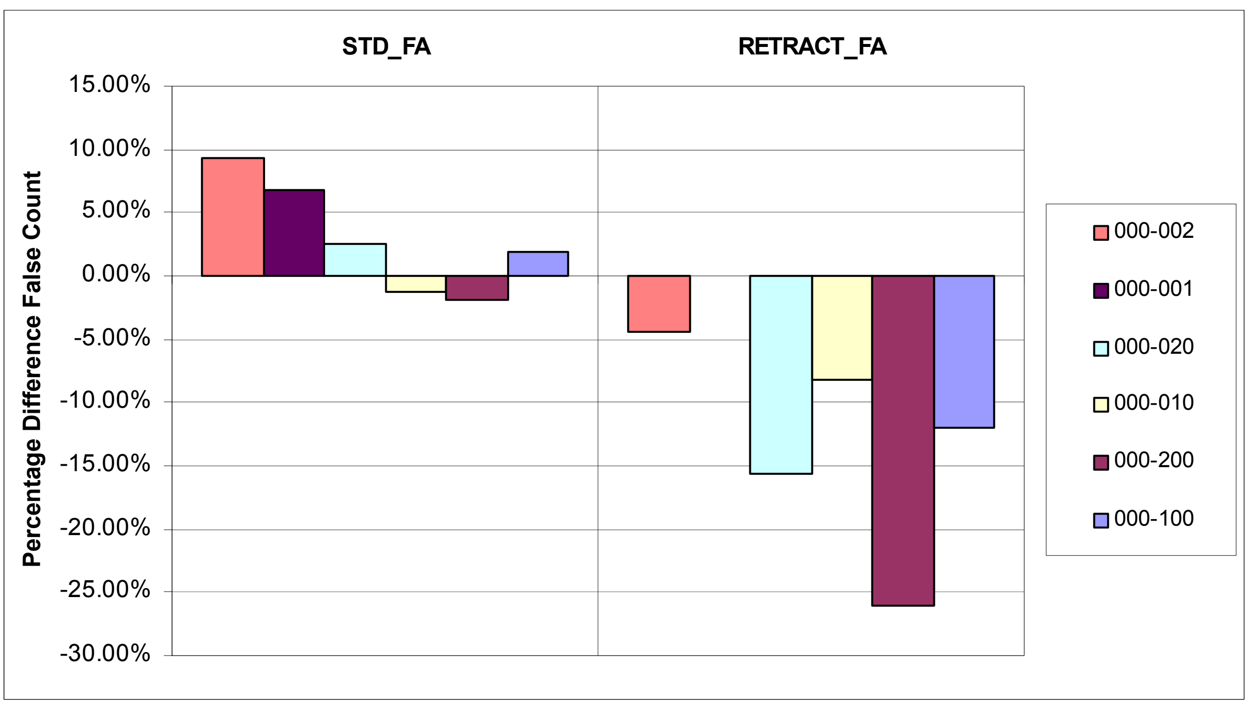


Figure 3.2-3: Analysis AB Run Percentage Difference of Standard and Retracted False Counts

3.2.2.4 Analysis BB's Run Comparison Conflict Prediction Results

The following Table 3.2-18 illustrates the results of a conflict prediction accuracy comparison under the Analysis BB which considers both red and yellow alerts and ignores adherence. Once again, the evaluations codes presented in this table are defined in the methodology Section 2.2.3.2.2. The difference in missed alert probabilities range from -0.04 to 0.01 without any specific pattern between runs. The false alert probability difference between the control and treatment runs ranges from -0.03 to 0.01. These false alert probabilities are higher for the wind treatment runs 100, 200, 010, and 020 only.

Table 3.2-18: Conflict Prediction Run Comparison Statistics for Analysis BB

Comparative Counts		Comparison Runs: Run A Versus Run B					
Event	Evaluation Code	000-100	000-200	000-010	000-020	000-001	000-002
V_{A1} or V_{B1}	SAME_VA	165	162	166	167	171	171
M_{A1} or M_{B1}	SAME_MA	31	31	33	30	35	32
F_{A1} or F_{B1}	SAME_FA	376	355	375	344	383	349
V_{A2} or M_{B2}	VA_MA	9	12	8	7	3	3
M_{A2} or V_{B2}	MA_VA	6	6	4	7	2	5
V_{A3} or $Discard_B$	VA_DISCARD	0	0	0	0	0	0
$Discard_A$ or V_{B3}	DISCARD_VA	0	0	0	0	0	0
F_{A2} or NC_B	FA_NC	73	94	74	105	66	100
NC_A or F_{B2}	NC_FA	109	167	100	139	57	80
Comparative Statistics		000-100	000-200	000-010	000-020		000-002
$P(M_A\|C_A)$		0.175	0.175	0.175	0.175	0.175	0.175
C_A		211	211	211	211	211	211
$P(M_B\|C_B)$		0.190	0.204	0.194	0.175	0.180	0.166
C_B		211	211	211	211	211	211
$P(M_A\|C_A) - P(M_B\|C_B)$ [Eq. 6]		-0.014	-0.028	-0.019	0.000	-0.005	0.009
$P(F_A\|A_A)$		0.721	0.721	0.721	0.721	0.721	0.721
A_A		623	623	623	623	623	623
$P(F_B\|A_B)$		0.739	0.757	0.736	0.735	0.718	0.709
A_B		656	690	645	657	613	605
$P(F_A\|A_A) - P(F_B\|A_B)$ [Eq. 10]		-0.019	-0.036	-0.016	-0.014	0.003	0.012
C_{ALL} [Eq. 3]		211	211	211	211	211	211
A_{ALL} [Eq. 13]		738	796	727	769	682	708
$(M_{A1}$ or $M_{B1})/ C_{ALL}$ [Eq. 14]		0.147	0.147	0.156	0.142	0.166	0.152
$(V_{A1}$ or $V_{B1})/ C_{ALL}$ [Eq. 15]		0.782	0.768	0.787	0.791	0.810	0.810
$(V_{A2}$ or $M_{B2})/ C_{ALL}$ [Eq. 16]		0.043	0.057	0.038	0.033	0.014	0.014
$(V_{B2}$ or $M_{A2})/ C_{ALL}$ [Eq. 17]		0.028	0.028	0.019	0.033	0.009	0.024
$(F_{A1}$ or $F_{B1})/ A_{ALL}$ [Eq. 18]		0.509	0.446	0.516	0.447	0.562	0.493

As defined in Section 2.2.3.2.1 and discussed in the previous Section 3.2.2.1, the statistical hypothesis tests applied are presented in the Table 3.2-19 and Table 3.2-20. Table 3.2-19 provides statistical evidence that the false alert probabilities are higher for the wind treatment runs 100, 200, 010, and 020, and not higher for the air temperature runs. Table 3.2-20 provides statistical evidence that the missed alert events probabilities are equivalent between the control run and all the treatment runs. In other words, for Analysis BB the wind treatment runs has a statistically significant effect on the false alert probability. However, the actual effect was not much in terms of false alert probability with the largest increase of 0.04 for Run 200. Expressed as a percentage of the control run's false alert count, the difference in the false alert counts is illustrated in Figure 3.2-4, which shows the retracted false alerts responsible for the increase.

Table 3.2-19: Analysis BB False Alert Event Statistical Test

	Comparison Runs: Run A Versus Run B					
Statistics	000-100	000-200	000-010	000-020	000-001	000-002
FA_NC	73	94	74	105	66	100
NC_FA	109	167	100	139	57	80
Total	182	261	174	244	123	180
Expected	91	130.5	87	122	61.5	90
X^2	7.121	20.418	3.885	4.738	0.659	2.222
P-value	0.008	0.000	0.049	0.030	0.417	0.136

Table 3.2-20: Analysis BB Missed Alert Event Statistical Test

	Comparison Runs: Run A Versus Run B					
Statistics	000-100	000-200	000-010	000-020	000-001	000-002
VA_MA	9	12	8	7	3	3
MA_VA	6	6	4	7	2	5
Total	15	18	12	14	5	8
Expected	7.5	9	6	7	2.5	4
X^2	0.600	2.000	1.333	0.000	0.200	0.500
P-value	0.439	0.157	0.248	1.000	0.655	0.480

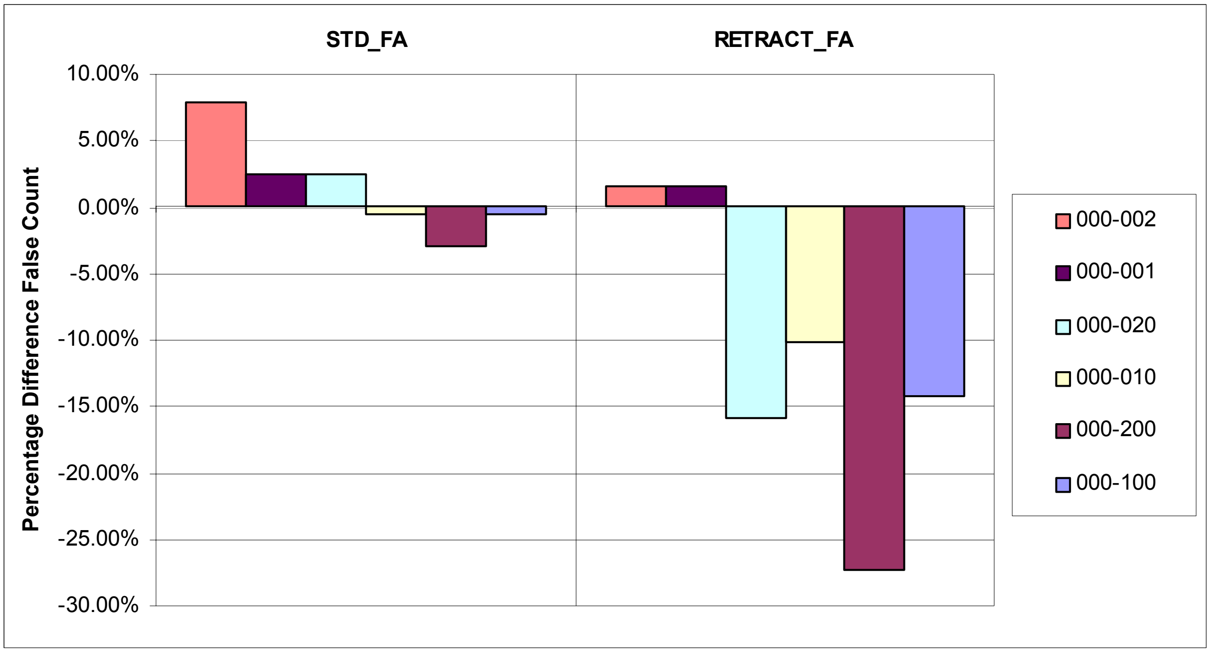


Figure 3.2-4: Analysis BB Run Percentage Difference of Standard and Retracted False Counts

3.2.3 Results on Comparison of Common Valid Alert Attributes

As described in detail in Section 2.2.3.3, the common valid alerts are now examined in terms of warning time and predicted conflict start time. For this analysis, ignoring or requiring adherence makes no difference, since adherence only influences conflict prediction errors, such as missed and false alerts. This section will only present results on the Analyses AA and BB, since the other two analyses have identical results to these.

Table 3.2-21 provides detailed results on comparing the warning time of common valid alerts per run. For Analysis AA, only the high level treatment runs had a statistically significant effect on warning time (i.e. 200, 020, and 002). However, all the runs had a median value of zero (i.e. the 50th quantile). Fifty percent of the data as expressed by the range between the 75th and 25th quantiles is also zero. The 200, 020, and 002 runs have an average warning time deviation from the control run of 14, 10, and 11 seconds, respectively. Therefore, all have no practical significance. The average warning time deviation would have to be at least 30 seconds or more to have any practical significance.

Table 3.2-21 provides results comparing warning time of common valid alerts for the Analysis BB also. Using a 0.10 confidence level, the effect is even less than Analysis AA with none of the run comparisons having any statistical significance. Therefore, for all run comparisons the study provides no evidence that the induced weather forecast errors have an effect on the warning time.

Table 3.2-21: Comparison of Common Valid Alert's Warning Time

Run A Vs. Run B	Wilcoxon Signed-Rank Test (P-value)	10 Quantile	25th Quantile (sec)	Quantile (sec)	75th Quantile (sec)	90th Quantile (sec)	Sample Mean (sec)	Sample Standard Deviation (sec)	Practical Significance
000-100AA	0.436	-11	0	0	0	22	5.5	78.8	N
000-200AA	0.006	-1	0	0	0	99	14.2	80.6	N
000-010AA	0.281	-11	0	0	0	36	5.9	76.5	N
000-020AA	0.083	-15	0	0	0	51	10.2	68.4	N
000-001AA	0.930	0	0	0	0	0	-3.6	38.1	N
000-002AA	0.020	0	0	0	0	24	10.8	65.5	N
000-100BB	0.465	0	0	0	0	12	0.1	51.7	N
000-200BB	0.761	-26	0	0	0	25	0.2	71.1	N
000-010BB	0.808	-5	0	0	0	0	-5.0	74.2	N
000-020BB	0.111	0	0	0	0	32	7.2	78.2	N
000-001BB	0.638	0	0	0	0	0	-1.4	40.5	N
000-002BB	0.306	0	0	0	0	12	1.9	39.8	N

Next, the common valid alerts were compared in terms of predicted conflict start time. Table 3.2-22 Shows that none of the Analysis AA results have a statistically significant difference in predicted conflict start time. For Analysis BB, only Run 010 had a statistically significant difference, but the average difference was only about three seconds. Therefore, the study does not provide evidence to state that the predicted conflict start time is different between the control run and the treatment runs.

Table 3.2-22: Comparison of Common Valid Alert's Predicted Conflict Start Time

Run A Vs. Run B	Wilcoxon Signed-Rank Test (P-value)	10th Quantile (sec)	25th Quantile (sec)	50th Quantile (sec)	75th Quantile (sec)	90th Quantile (sec)	Sample Mean	Sample Standard Deviation (sec)	Practical Significance
000-100AA	0.895	-16	-4	0	3	10.4	-1.2	30.8	N
000-200AA	0.734	-18	-7	0	5	17	-0.5	37.7	N
000-010AA	0.377	-16	-5	0	3.25	16.5	-0.2	19.3	N
000-020AA	0.17	-21.4	-8	0	4	15.8	-0.1	24.2	N
000-001AA	0.643	-3	-1	0	1	5	-0.5	13.7	N
000-002AA	0.158	-15	-3.25	0	1.25	12.7	-0.4	20	N
000-100BB	0.647	-17	-4	0	5	12	-1.9	28.6	N
000-200BB	0.855	-18	-7	0	6	27.4	3.4	43.5	N
000-010BB	0.046	-18	-6	0	3	10	-2.8	20.1	N
000-020BB	0.11	-29	-9	0	5	13	-3.8	25.6	N
000-001BB	0.193	-3	-1	0	2	5.8	0.3	15.3	N
000-002BB	0.232	-13.6	-4	0	2	15.8	1	24.4	N

3.2.4 Summary of Conflict Prediction Accuracy Results

The conflict prediction analysis performed in this study and reported in Section 3.2 was partitioned into three subsections. First, Section 3.2.1 reported on the individual run results. It reported the missed, valid, and false alert counts and associated probabilities for each of the six treatment runs and one control run. Four analyses were performed for each of these seven runs. Red alert analysis with and without considering flight plan adherence was reported. All alerts (both red and yellow) with and without considering flight plan adherence were examined as well. Thus, the analysis was very comprehensive in considering both alert type and flight plan adherence, but the study also provided the specific reasons for the missed and false alert designations. For example, false alerts were subdivided into retracted or standard false alerts and missed alerts were subdivided into no-call and late missed alerts.

Next, in Section 3.2.2 the individual treatment runs were each compared to the control run. The paired run comparisons included all combination of error events for each of the four analyses. For example, if both the control and treatment run correctly predicted a conflict, the count was labeled SAME_VA, but if both runs incorrectly predicted the conflict, it was labeled SAME_MA. However, the most interesting analysis result was not when these run comparisons demonstrate the treatment agreeing with the control run, but the opposite result when they were in disagreement. In regards to missed alert error, this occurred as VA_MA and MA_VA counts. A significant difference between these counts would provide evidence to state that the control and treatments runs did have a different missed alert probability. For all four analyses and for all control and treatment run comparisons, these counts were compared and none were found to be significantly different in this experiment. An analogous analysis was performed with the false alert error. These events were labeled and counted as NC_FA and FA_NC. Once again, the "NC" portion of the label designates that URET correctly did not present an alert (no-call) for an encounter between two aircraft not violating separation standards. The "FA" portion of the label designates a false alert call where the predicted conflict notification is presented. A NC_FA count refers to the control run having a "NC" event and the treatment a "FA" event, while the FA_NC is the opposite outcome. If these two counts were significantly different, the test would provide evidence to state the false alert probability was different between the control and treatment runs (e.g. higher or lower). Fairly consistently, in all four analyses for all run comparisons, the wind magnitude and wind direction did have a significant difference, but the air temperature runs consistently did not. However, reviewing the effect of the wind magnitude and wind direction comparisons to the control run indicates they may not have practical significance. For all analyses and runs, the highest difference in false alert probability between the control and treatment run comparisons was about 0.06, which occurred in analysis AA with run comparison 000-020. For all runs, the false alert probability difference ranged from 0.01 to 0.06. Furthermore, the difference was dominated by the retracted false alert events. This was illustrated in Figure 3.2-1, Figure 3.2-2, Figure 3.2-3, and Figure 3.2-4 that plotted the percentage in false alert differences for both standard and retracted false alerts for all for analyses, respectively.

Finally, in Section 3.2.3 the common valid alerts for each run comparison were examined by comparing conflict warning time and predicted conflict start time errors. In summary, only two run comparisons had a warning time with a significant difference (i.e. a P-value less than 0.10), but the highest mean different was only 14 seconds. Only one run comparison had a predicted conflict start time error with a significant different, but it was only 3 seconds different on average. Therefore, for both the warning time and predicted conflict start time error no practical significance was evaluated, assuming at least 30 seconds mean difference would be required.

3.3 Flight Examples

Statistical results have just been presented for the effects of the introduced weather forecast errors on the entire scenario. This section presents examples of the experimental results on individual flights selected from the various runs. Details of the effects of the errors on specific flights are presented. The emphasis is on illustrating the difference between the performance from the control run (i.e. Run 000) and one or more of the treatment runs (e.g. Run 100, 200, 010, 020, etc.).

A detailed examination of the first flight example is included in Appendix D. This examination of a typical flight shows how the various induced errors in the treatment runs affected trajectory and conflict prediction accuracy. A brief description of the effects of wind magnitude, resulting from Run's 100 and 200, is given in the following Section 3.3.1. The next three flight examples presented in Sections 3.3.2, 3.3.3, and 3.3.4 focus on the impact on conflict prediction accuracy.

3.3.1 Flight Example #1

The flight selected for detailed analysis was an Embraer Brasilia, a twin engined turboprop, flying from Cincinnati, OH to Madison, WI via Muncie, IN and Northbrook, IL at a cleared altitude of 20,000 feet. There were no flight plan amendments.

3.3.1.1 Flight Path

The track of the aircraft in the ZID weather scenario is a climb out of Cincinnati, starting at 6750 feet, climbing to 20,000 feet. The aircraft deviates to the right of the filed route, flying a heading of 335 degrees, causing URET to update its predicted trajectories several times during the climb. At 18,000 feet, the aircraft starts turning to a heading of 295 degrees and the predicted trajectory coincides laterally with the future track.

The path of the flight in the scenario is shown in Figure 3.3-1 as the heavy weight black line. The distances in the plot are the Center's stereographic coordinates. The flight starts in the Indianapolis Center (ZID) airspace and continues into the Chicago Center (ZAU) airspace. The border between the ZID and ZAU airspaces is the medium weight black line. The trajectories generated by the URET conflict probe are the light weight lines. The starting point for each trajectory is the location of the aircraft when the prediction is made. URET generated seven trajectories for this sample flight. The first and second and the third and fourth trajectories predict the same route – each pair forms a single trace in Figure 3.3-1. This only five distinct trajectories can be seen on the plot. The trajectory traces are numbered – 1 through 7 – in the time order that they were generated.

The vertical profiles of the track and the trajectories are plotted in Figure 3.3-2. The vertical profile of a flight is a plot of the altitude of the flight versus time. The heavy trace is the profile the aircraft actually flew. The light weight lines are the profiles or trajectories that URET predicted. The plot shows that the aircraft in the scenario started at about 6000 feet and climbed to 20,000 feet where it leveled off.

3.3.1.2 Forecasted Wind Runs

The magnitude of the (horizontal) wind along the path of flight for the control run is plotted in Figure 3.3-3 as the top trace. The abscissa in the plot is time, which directly corresponds to the movement of the aircraft along its route. The relative wind – the component of the wind directly opposing the forward progress of the aircraft – is plotted for the control run, for the plus 20 knots run, and for the plus 60 knot run. The relative winds are shown as negative because they are headwinds. The magnitude and direction of the wind change slowly along the aircraft's track – the changes in the relative wind are caused mainly by the changes in the aircraft's heading. At approximately 70400 seconds, the headwind decreases substantially for a short period of time – from –21 knots to –9 knots. The reduction is accentuated by the wind magnitude errors. In Run 200 (the 60 knot error run) the headwind is reduced briefly from –70 knots to –25 knots.

The top trace in Figure 3.3-4 gives the direction of the aircraft's track. The bottom two traces give the wind direction and the direction of the wind relative to the aircraft track respectively. The relative wind direction is the difference between the direction of the track and the direction of the wind. The directions are illustrated as functions of the movement of the aircraft along its path of flight.

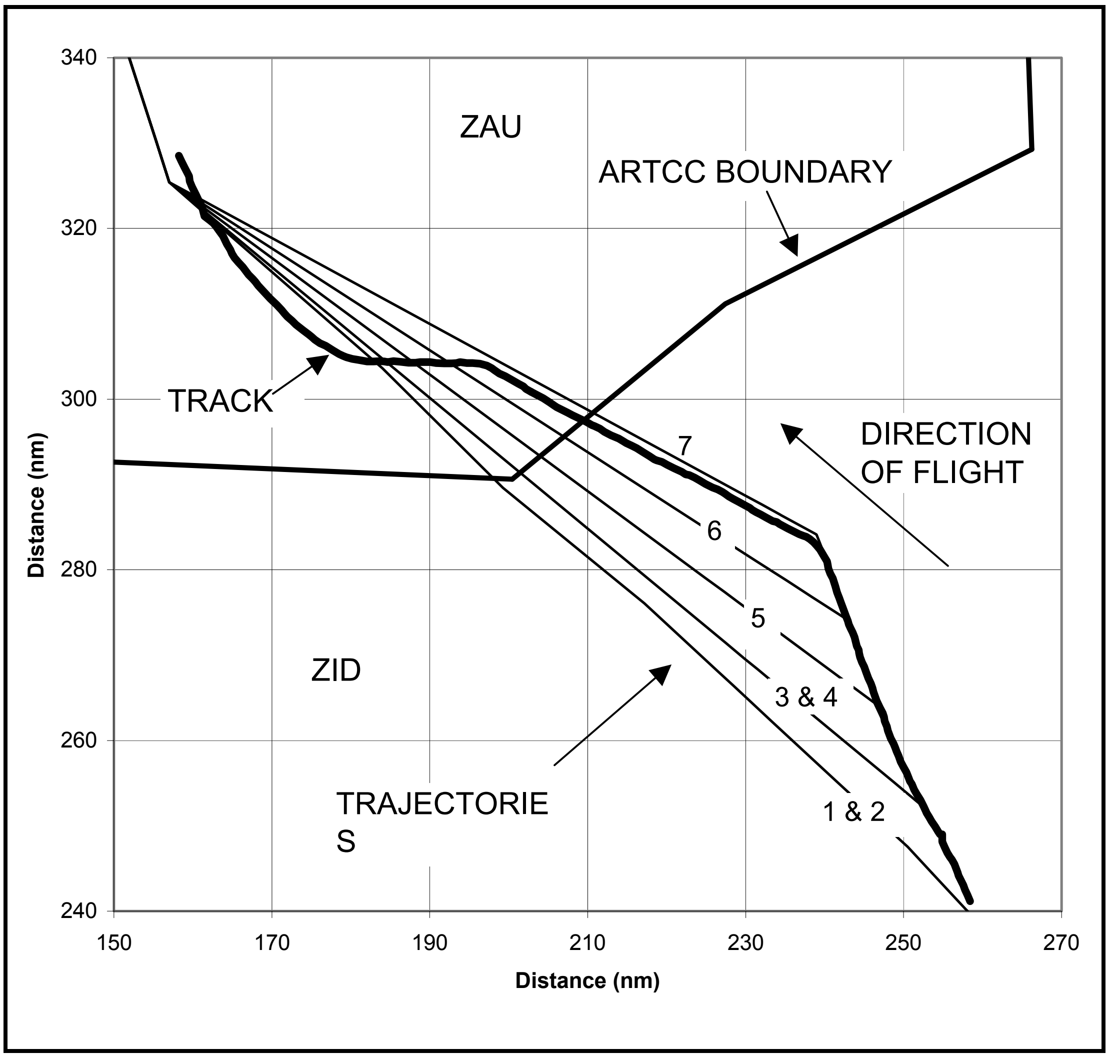


Figure 3.3-1 Example 1 Route

The wind direction trace gives the wind direction for each point on the path the aircraft flies. It shows that the wind direction changes only moderately during the flight. The changes in the relative wind direction therefore mimic the changes in the direction of the aircraft track. The big change in the relative wind direction at around 70400 seconds is reflected in the big change in the relative wind magnitude shown at 70400 seconds in Figure 3.3-3.

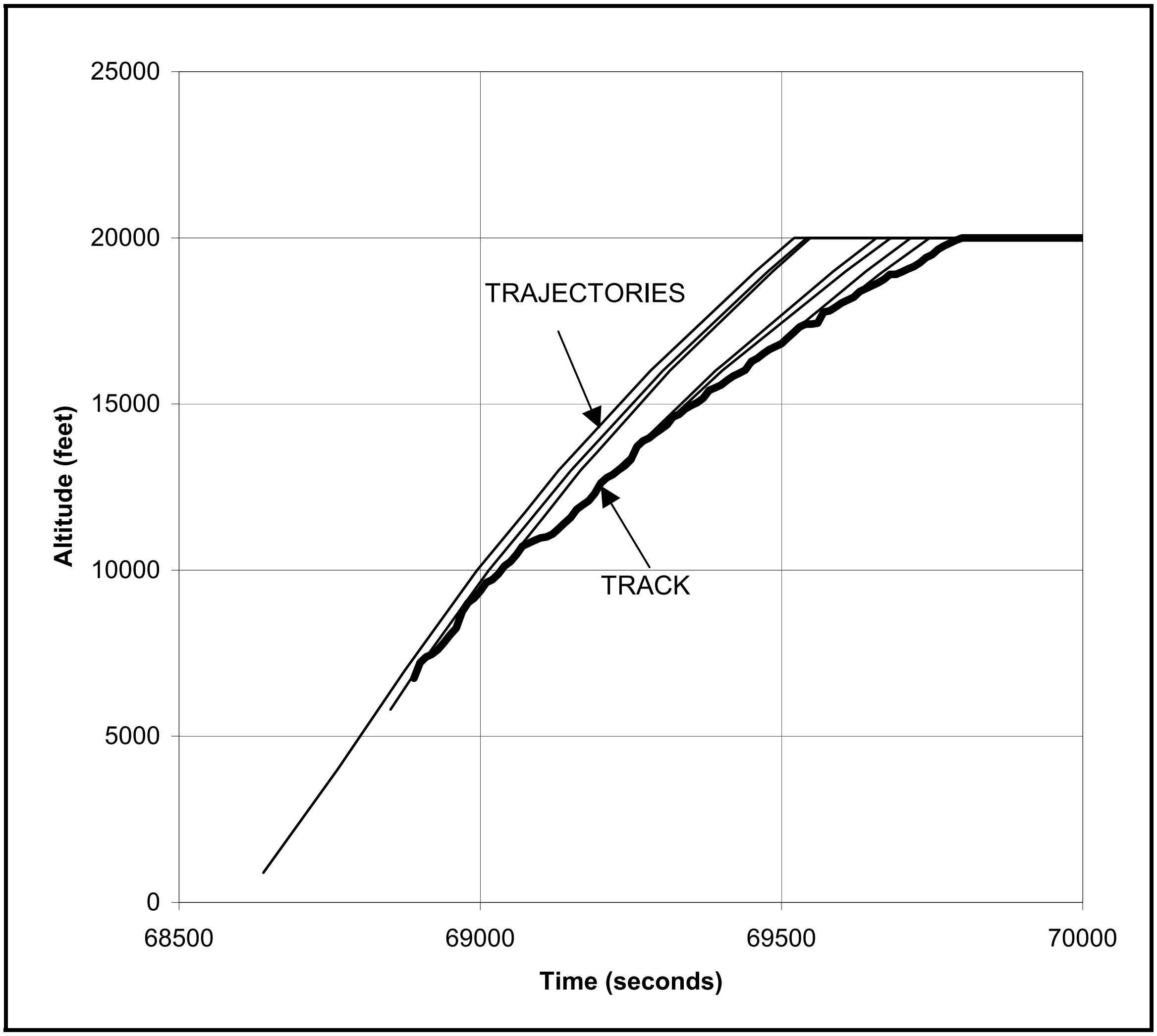


Figure 3.3-2 Example 1 Vertical Profiles

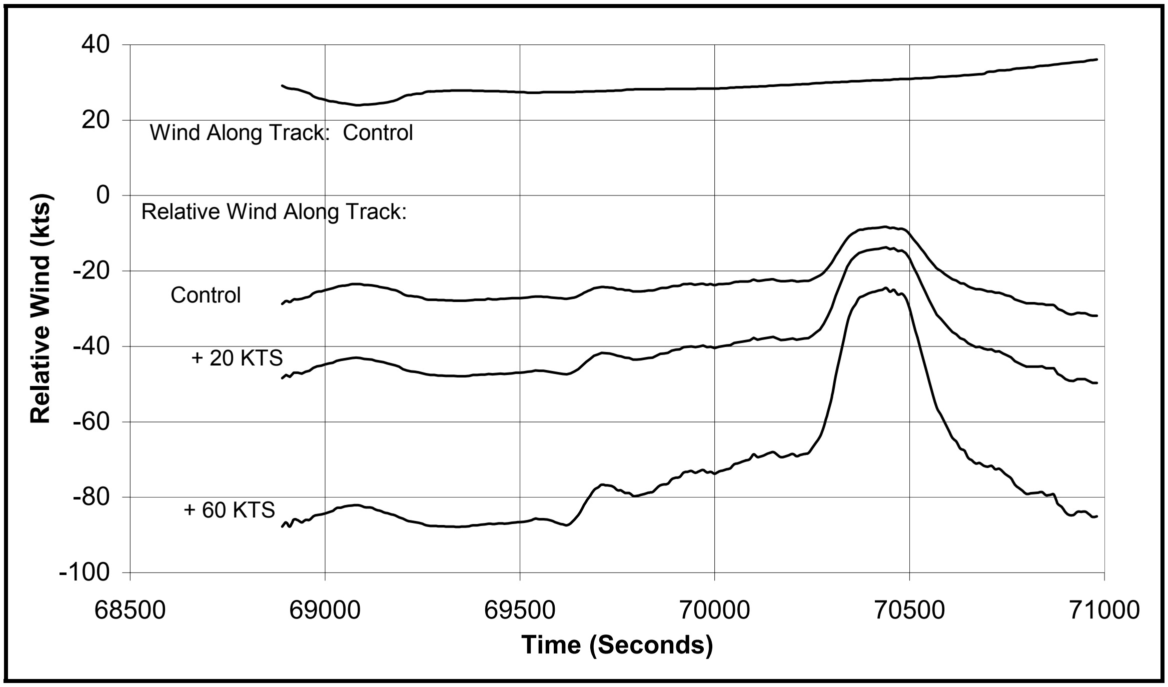


Figure 3.3-3 Example 1 Relative Wind Along Track

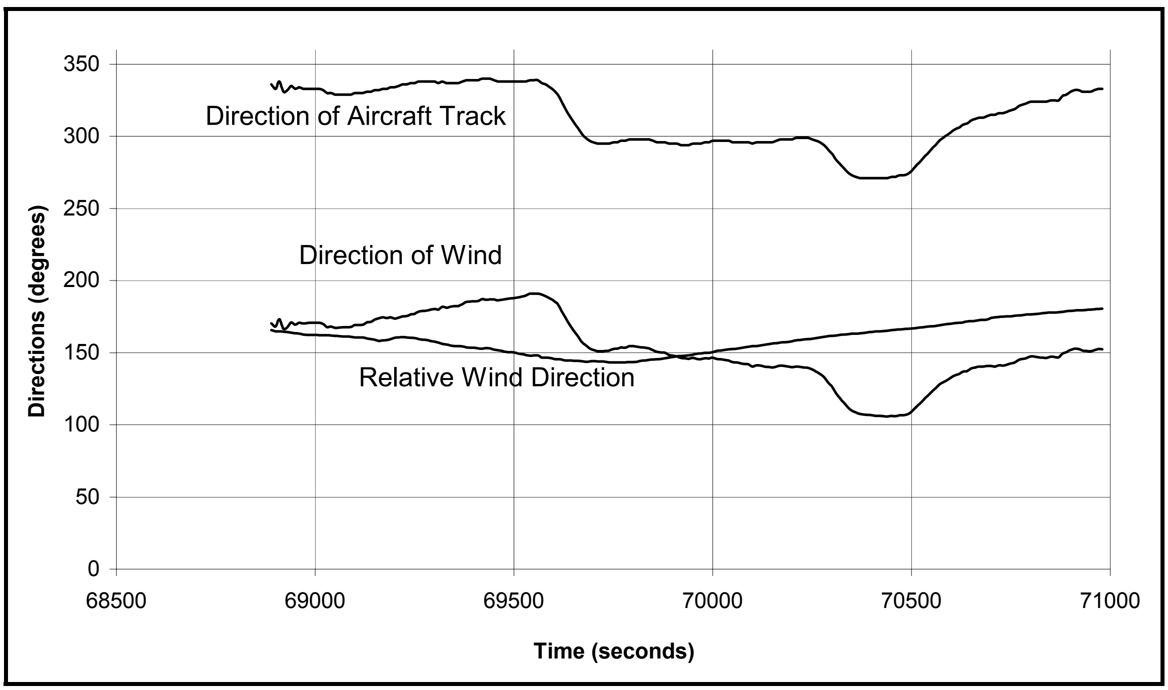


Figure 3.3-4 Example 1 Track and Wind Directions Along Track

3.3.1.3 Trajectories

As the aircraft flies along its route, URET updates its prediction of the future flight path of the aircraft. Initially the aircraft does not follow its filed flight plan route and URET is forced to update its predictions repeatedly. In the control run URET builds seven trajectories during the time the flight is in the scenario. These trajectories are shown with the aircraft track in Figure 3.3-1. The system consistently predicts a return to a downstream fix. The plots of the vertical profiles of the trajectories in Figure 3.3-2 show that URET over estimated the rate of climb. The slopes of the traces in Figure 3.3-2 are a measure of the rates of climb of the aircraft.

In the second run, Run 100, with an introduced wind magnitude error of plus 20 knots, one more trajectory is generated by URET, making a total of eight; in the third run, Run 200, with an introduced wind magnitude error of plus 60 knots, three additional trajectories are generated for a total of ten. The errors in the predicted winds in the second and third scenarios cause the track to deviate from the predicted trajectories more rapidly than in the control run. The longitudinal errors in aircraft position are increased. The lateral and vertical errors are mostly unaffected. URET responds by updating the trajectories more frequently.

3.3.1.4 Ground Speed

In Runs 100 and 200 the headwinds are predicted to be larger than they actually are when the aircraft flies its route. The aircraft is predicted to fly more slowly than it actually does. This effect is illustrated in the following figures. In Figure 3.3-5 and Figure 3.3-6, the predicted along track distances (the trajectories) are plotted versus time along with the actual distance flown (the track). The slopes of the lines are the predicted speeds of the aircraft. The trajectories, the straight lines in the plot, predict the distance traveled. They are labeled with the times the predictions were made. The vertical lines indicate the times when the trajectories were built. Only the first part of each trajectory is plotted.

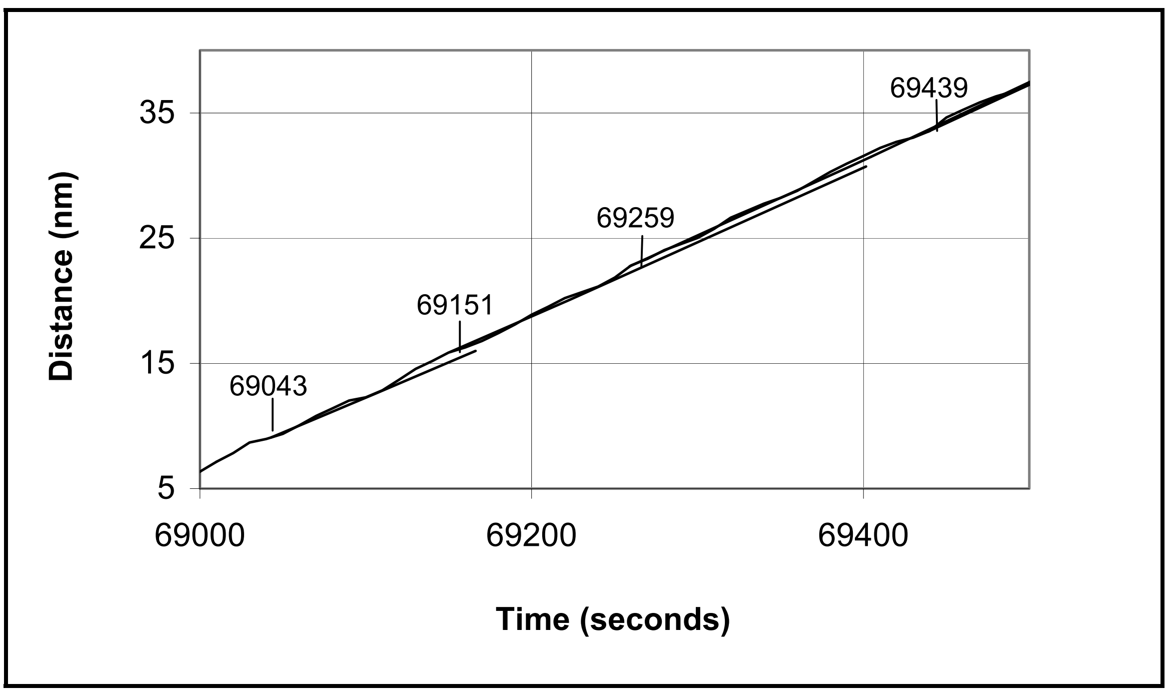


Figure 3.3-5 Example 1 Cumulative Along Track and Trajectory Distances – Control Run

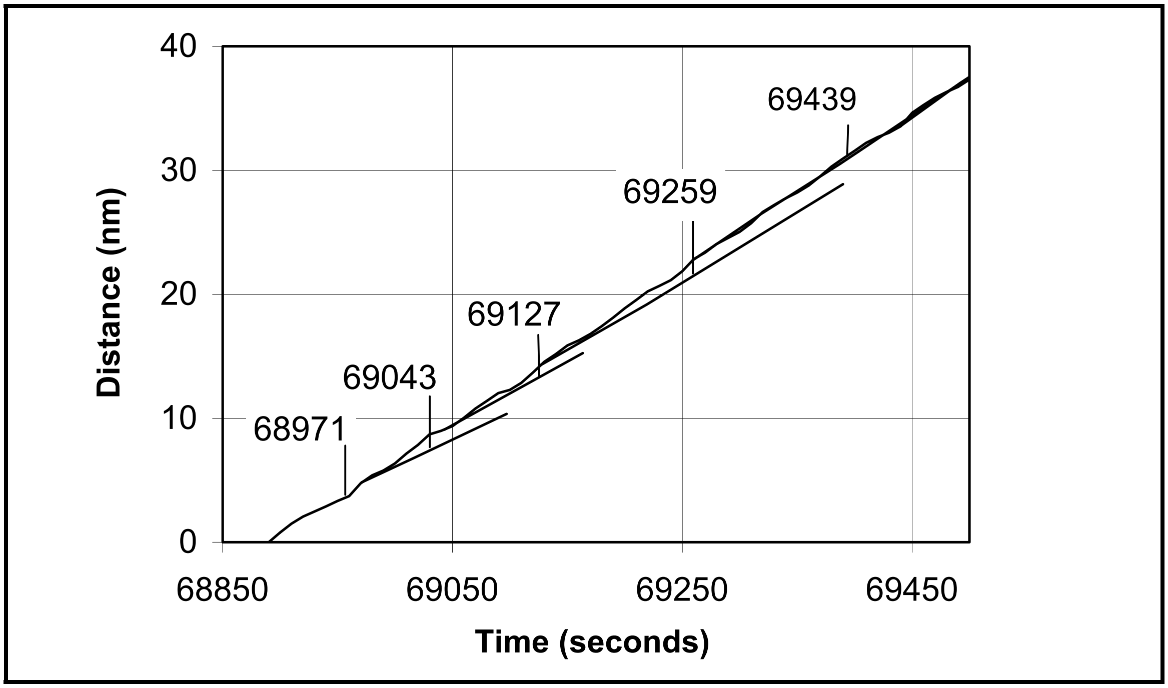


Figure 3.3-6 Example 1 Cumulative Along Track and Trajectory Distances – Run 200

In the control run of Figure 3.3-5, the wind and speed predictions are accurate. As a result, the plot of the distance accumulated by the track, lies on top of the trajectory plots. In Run 200 (plus 60 knot run), the trajectories are in error. A larger headwind was predicted than actually occurred. The plots of the trajectories in Figure 3.3-6 are consistently below the track. They are predicting a slower ground speed for the aircraft than actually occurred.

These figures also illustrate that the URET system responds to the wind error by building new trajectories at different times. Note that the trajectories that are built at the same time in the two different runs are not the same; they predict different speeds.

3.3.1.5 Climb Angle

The errors in the predicted headwinds cause URET to predict steeper climbs than actually occur. The three runs are compared in Figure 3.3-7 using the three trajectories built at time 69043 seconds UTC. The altitude versus distance is plotted for each trajectory. The slopes of the trajectories (and also of the track) give the angle of climb. The increased predicted headwinds cause the plus 20-knot trajectory to be steeper than the control trajectory and the plus 60-knot trajectory to be steeper than the plus 20-knot trajectory.

3.3.1.6 Climb Rate

The predicted rate of climb is unaffected by errors in the predicted horizontal winds. A plot of altitude versus time for the three trajectories built at 69043 seconds UTC in the three runs in Figure 3.3-8 illustrates this. The rates of climb are the slopes of the traces. The slopes of the three trajectories are almost identical, showing that the predicted climb rate is practically the same for all three runs.

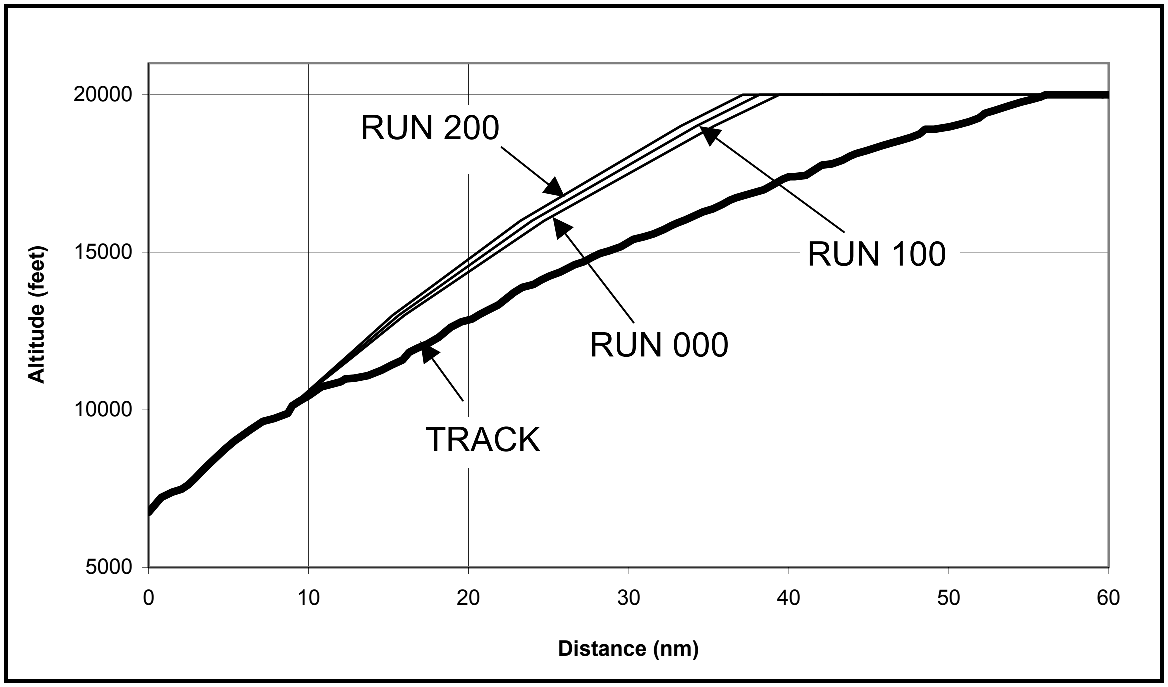


Figure 3.3-7 Example 1 Angle of Climb - Trajectories Built at 69043

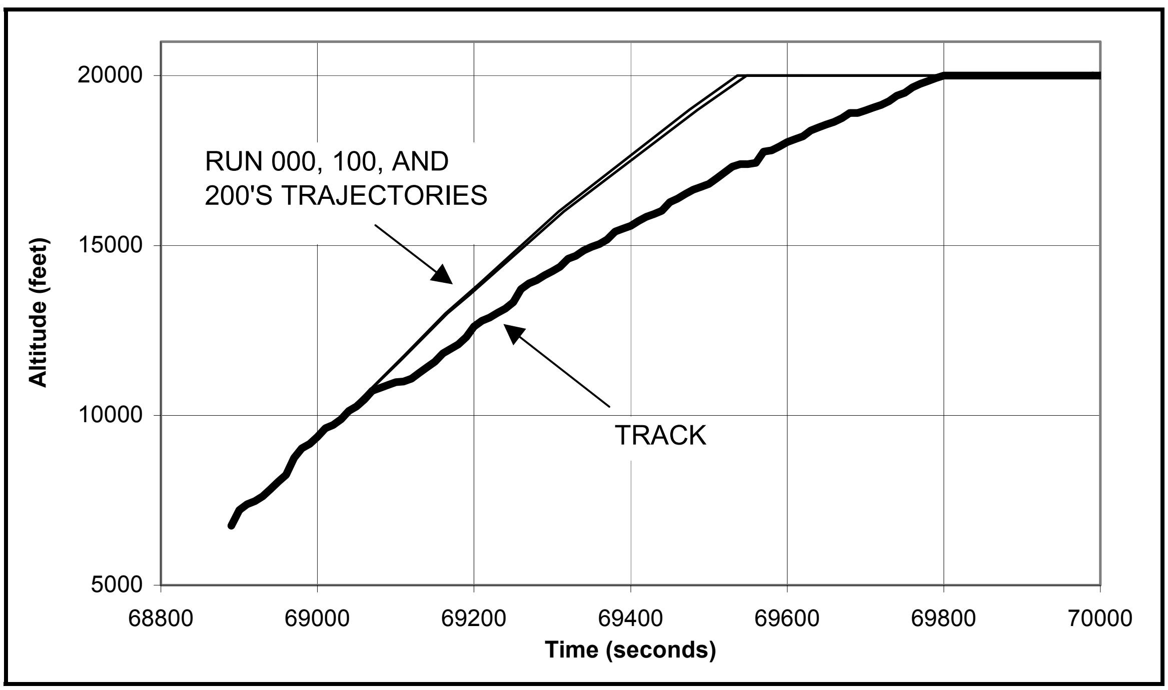


Figure 3.3-8 Example 1 Rate of Climb - Trajectories Built at 69043

3.3.1.7 Forecasted Wind Direction Runs

Two runs were made with errors introduced into the forecasted wind directions. These errors have the effect of adding errors to the forecast wind magnitudes. The results are similar to the wind magnitude runs as described in Section 3.3.1.3. Details are given in Appendix D.

3.3.1.8 Forecasted Air Temperature Runs

Two runs were made with errors added to the forecast air temperatures. The main effect is to cause URET to predict a reduced rate of climb. This is turn causes an increase in errors in the predicted altitudes of the aircraft. Details are given in Appendix D.

3.3.1.9 Conflict Predictions

The conflict predictions made by URET are based on the trajectories. The errors introduced into the weather data cause changes in the trajectories, which in turn cause changes in the conflict predictions. The conflict predictions are analyzed for this example using the Analysis BB. In this analysis, both red and yellow alerts are counted and lack of adherence to professed intent is ignored.

The flight has no conflicts but it does have encounters with six other aircraft. The minimum horizontal separations (independent of vertical separation) range from 3.3 nm to 23.5 nm.

The control run has three notification sets predicting crossing conflicts with three different aircraft. These aircraft are designated here as WXY1000, WXY2000, and WXY3000. The first alert is retracted; the other two are not. These notification sets are listed as the first three rows in Table 3.3-1. The entry in the first column of the first row is labeled “000BB-WXY1000.” The “000” identifies the run as the control run; the “BB” specifies the analysis method (i.e. adherence ignored with both red and yellow alerts counted). WXY1000 identifies the aircraft predicted to be in conflict with the example flight, ABC1000.

Table 3.3-1 Example 1 Notification Sets for ABC1000 with Other Flights

Run and Aircraft ID of Other Flights	Notification Set Start Time	Notification Set End Time	Predicted Conflict Start Time	Predicted Conflict End Time	Description
000BB-WXY1000	19:05:17	19:08:06	19:16:10	19:17:14	Retracted false alert
000BB-WXY2000	19:15:34	19:23:58	19:21:23	19:23:02	Standard false alert
000BB-WXY3000	19:22:14	19:24:42	19:24:10	19:24:41	Standard false alert
100BB-WXY1000	19:05:08	19:08:06	19:16:17	19:17:35	Retracted false alert
100BB-WXY2000	19:15:34	19:23:57	19:21:24	19:22:57	Standard false alert
100BB-WXY3000	19:24:32	19:24:51	19:24:32	19:24:50	Standard false alert
100BB-WXY4000	19:06:00	19:07:31	19:12:08	19:12:29	Retracted false alert
100BB-WXY5000	19:17:19	19:18:50	19:24:42		Retracted false alert
200BB-WXY1000	19:04:27	19:08:06	19:15:53	19:17:30	Retracted false alert
200BB-WXY2000	19:15:34	19:23:59	19:21:24	19:22:43	Standard false alert
200BB-WXY3000	19:22:14	19:25:06	19:24:44	19:25:12	Standard false alert
200BB-WXY4000	19:06:00	19:08:52	19:12:04	19:13:08	Retracted false alert
200BB-WXY5000	19:17:19	19:18:51	19:24:42	19:24:58	Retracted false alert
200BB-WXY6000	19:07:40	19:09:31	19:13:19	19:13:45	Retracted false alert

The notification sets for Run 100 are listed in the table in rows 4 through 8, for Run 200 in rows 9 through 14. In Run 100 and Run 200, the three false alerts found in Run 000 are repeated with small changes in the times. In Run 100, and two new false alerts are added. These two new alerts are retracted. In Run 200, three additional false alerts are added. They are retracted also.

Run 000 predicts three conflicts, Run 100 predicts five conflicts, and Run 200 predicts six conflicts. The aircraft predicted to be in conflict with ABC1000 are identified in column one of the table as flights WXY1000 through WXY6000. The alerts in this example are all false; the majority are retracted.

3.3.1.10 Summary of Flight Example #1

The flight example has shown that adding error to the forecast winds aloft causes additional errors in predicted positions. The new errors are principally along track or longitudinal errors caused by inaccurate predictions of ground speed. These errors are increased by increases in the errors in wind magnitude. The lateral errors, which are large due to lack of intent information, are not affected. Vertical position errors, which are due mainly to errors in climb rate, are affected only moderately by errors in predicted winds, which change the predicted climb angle. .

As expected, the weather-induced errors in the predicted positions cause additional errors in the conflict predictions. As URET reconforms or rebuilds its trajectories to correct for these errors in predicted position, it adds and retracts false alerts. Thus, more retracted false alerts are generated as illustrated in Table 3.3-1 and described in the previous Section 3.3.1.9.

3.3.2 Flight Example #2

This example illustrates the effect of errors in the forecasted wind directions on conflict predictions. The example selected is an encounter (not a conflict) between a flight referred to as DEF1000 and another aircraft designated as TUV1000.

3.3.2.1 Flight Path

The DEF1000 flight is an Embraer Brasilia, a twin engined turboprop, flying from Buffalo, NY to Cincinnati, OH via the Dryer VORTAC at Cleveland, OH and the Appleton VORTAC in OH. It flies at a cleared altitude of 22,000 feet using the CINCE STAR into the Cincinnati airport. The portion of the flight captured in the test scenario is its descent from 22,000 feet to 9,000 feet with a temporary level off at 11,000 feet. Interim altitude clearances for en route descent are issued for 18,000 feet, 17,000, 14,000, and 11,000 feet. This flight has an encounter with another aircraft, designated as TUV1000, but has no conflicts.

The TUV1000 flight is a Boeing 737-500 flying from Columbus, OH to Chicago Midway via the Rosewood VORTAC in OH and Fort Wayne, IN using the GOSHEN STAR into Midway airport. In the scenario, it climbs out of Columbus, starting at 3200 feet, climbing to 20,000 feet and later descending to 16,800 feet leveling off briefly at 18,000 feet during its descent.

3.3.2.2 Description of the Encounter

The encounter occurs when DEF1000 is descending into the Cincinnati airport and TUV1000 is climbing out of the Columbus airport. The minimum horizontal separation is 14.6 nm when the vertical separation is 3400 feet. The minimum vertical separation occurs 50 seconds later and is 200 feet when the horizontal separation is 18.1 nm.

3.3.2.3 Forecasted Wind Runs

DEF1000 has a slight tail wind at the first part of its flight. Later it turns into a head wind and remains a head wind until after its encounter with TUV1000. TUV1000 has a head wind for its entire scenario flight path. Comparing the control run (i.e. Run 000) with the larger wind magnitude run (i.e. Run 200) shows that the introduction of the 60 knot error in the winds increases the forecasted head wind on DEF1000 by about 25 knots immediately before the encounter and increases the forecasted head wind on TUV1000 by about 40 knots immediately before the encounter.

Each flight's root mean square (RMS) statistic of the differences between the control and treatment run's relative wind magnitudes is illustrated in Figure 3.3-9. It expresses the effects of the induced wind magnitude error on the relative winds for the two flights. The relative wind is the projected wind on each aircraft's path of flight, using its HCS track positions. The dark-shaded bars show this RMS statistic for DEF1000. The biggest change for DEF1000 occurs in the Run 020 when the forecasted wind direction is incremented by 90 degrees. Similarly the light-shaded bars in Figure 3.3-9 show the average change in the relative winds for TUV1000 along its flight path. The biggest change occurs with the Run 200, which includes the wind magnitude increase of 60 knots.

3.3.2.4 Forecasted Air Temperature Runs

The forecasted air temperatures are unchanged in Runs 100, 200, 010, and 020. They are altered in Runs 001 and 002 only. As expected the wind magnitude RMS statistic values in Figure 3.3-9 are zero for these runs. That is, the winds are unaltered in the air temperature runs (Run 001 and Run 002).

3.3.2.5 Longitudinal Errors

The changes in the predicted winds result in changes in the longitudinal errors in the trajectories that URET generates. The changes are illustrated in Figure 3.3-10. The longitudinal errors in each run are compared with the errors in the control run. The standard deviation for each comparison is shown for the two flights. The largest error increase is with DEF1000 in Run 200.

As shown in Figure 3.3-10, a comparison of the errors for the two flights for Runs 100 and 200 shows that the changes in predicted wind magnitude effect the longitudinal errors in DEF1000 more than TUV1000. Similarly, a comparison of the errors for the two flights for Runs 010 and 020 shows that the changes in predicted wind direction (which result in changes in wind magnitudes) effect the longitudinal errors in the TUV1000 flight more than the DEF1000 flight.

3.3.2.6 Conflict Predictions

In Run 200, URET predicts a conflict three times – in all cases with a yellow alert. Unlike Run 200, URET does not predict a conflict between DEF1000 and TUV1000 in the control run and other treatment runs. The conflict predictions are listed as notification sets in Table 3.3-2. When the track adherence rule is applied in the Analysis AB, two of the three alerts are discarded. The remaining alert is retracted before the predicted conflict start time. When the adherence rule is not applied in Analysis BB, all of the three yellow alerts are retracted before their predicted conflict start times.

3.3.2.7 Summary of Flight Example #2

The wind magnitude error induced in Run 200 as discussed above and illustrated in Figure 3.3-9 causes increases in longitudinal trajectory error. Once again, these trajectory errors are illustrated in Figure 3.3-10. The longitudinal trajectory error causes three yellow alert notification sets to be presented. In Analysis BB, these yellow alerts are all retracted false alerts. Therefore, for this example encounter, the wind magnitude error of Run 200 (i.e. adding 60 knots) causes an increase in retracted false alerts. This is consistent with Section 3.2's statistical analysis of all flights in the test scenario.

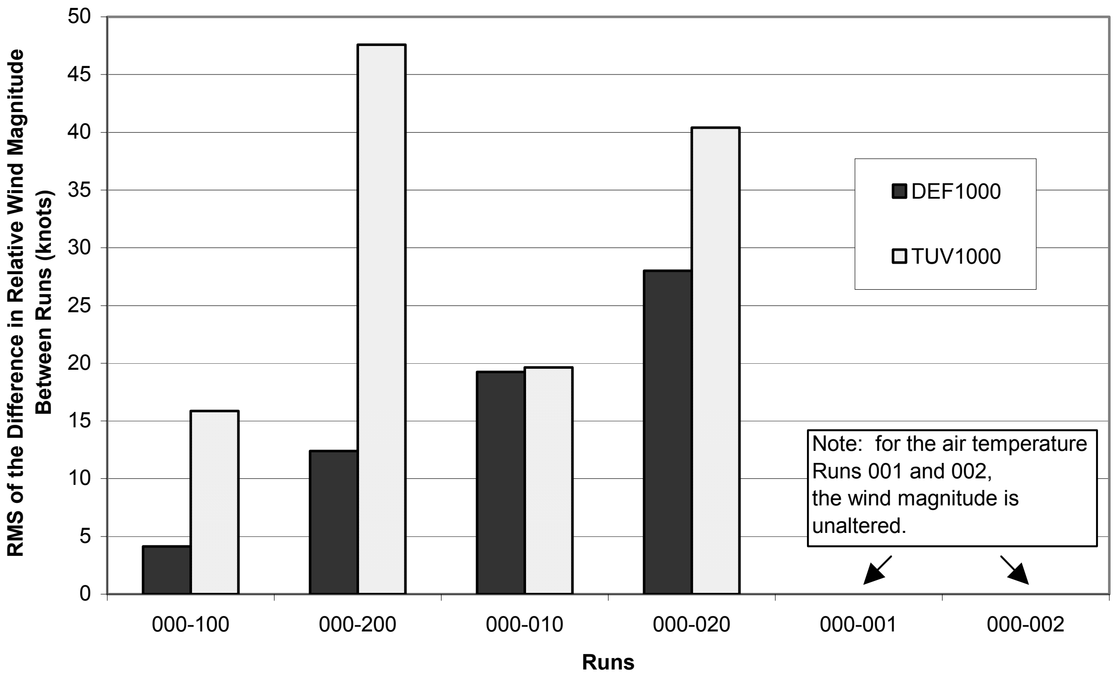


Figure 3.3-9 Example 2 RMS Values of the Differences in Relative Wind Magnitudes Between Runs

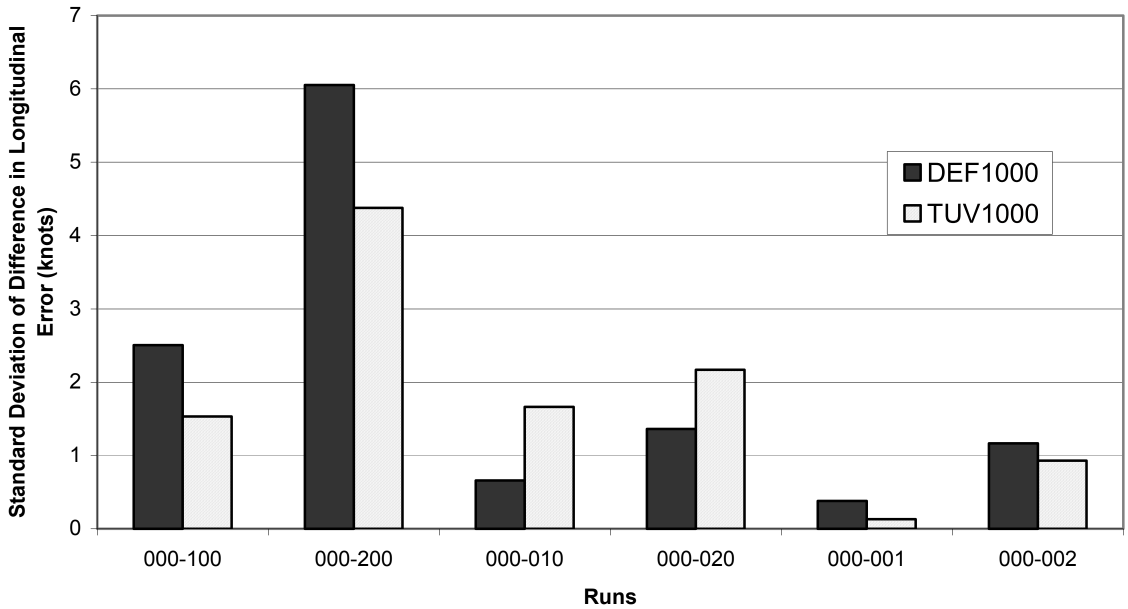


Figure 3.3-10 Example 2 Standard Deviations of the Differences in Longitudinal Errors between Runs

Table 3.3-2 Example 2 Notification Sets for Encounter with Flights DEF1000 and TUV1000

Run and Notification Set Number	Notification Set Start Time	Notification Set End Time	Predicted Conflict Start Time	Predicted Conflict End Time	
200AB-1	19:36:35	19:36:57	19:41:07	19:41:21	Retracted false alert
200AB-2	19:37:02	19:38:45	19:40:10	19:41:04	Discarded false alert due to out of adherence
200AB-3	19:39:23	19:39:33	19:39:46	19:40:10	Discarded false alert due to out of adherence
200BB-4	19:36:35	19:36:57	19:41:07	19:41:21	Retracted false alert
200BB-5	19:37:02	19:38:45	19:40:10	19:41:04	Retracted false alert
200BB-6	19:39:23	19:39:33	19:39:46	19:40:10	Retracted false alert

3.3.3 Flight Example #3

Flight Example #3 illustrates the effects of errors in forecasted wind direction on the performance of URET. It is an example of a level flight encounter between two aircraft, designated GHI1000 and QRS1000.

3.3.3.1 Flight Path

The GHI1000 flight is a Boeing 727-200 flying from Philadelphia to Dallas-Fort Worth at Flight Level (FL) 310. Initially its route is by Lancaster, PA and then on the J6 jetway to Little Rock. It is redirected to take the Henderson fix to Little Rock. From there it takes the Bonham 3 STAR into the Dallas-Fort Worth International Airport. The portion of the flight in the scenario is entirely at 31,000 feet.

The QRS1000 flight is a Boeing 737-400 flying from Philadelphia to Nashville at FL 310. Initially the flight is through Martinsburg, onto J6 jetway and then to the YOCKY fix. It is later redirected to take the DACOS fix to the GROAT fix, a fix downstream from YOCKY, and into Nashville via the GUITR fix. In the scenario, the flight is at FL 310 until 19:50:10 UTC when it starts its descent into Nashville, being cleared in steps to FL 240 and then FL 160.

3.3.3.2 Description of the Encounter

Both aircraft are cruising at FL 310 and about 19:11:00 UTC start in an in-trail encounter, separated about 16 nm horizontally. A little afterwards at 19:14:10 UTC, the aircraft come the closest with a minimum separation of 14.09 nm. At about 19:14:20 UTC, GHI1000 deviates in a northwest direction as discussed above (i.e. redirected to the Henderson fix). Later both flight paths are on parallel yet slightly converging routes for most of the remaining flight time. During this time, they converge from over 25 nm to about 15 nm and then diverge again when QRS1000 is redirected in a southwest direction (i.e. redirected to the DACOS fix, etc.).

3.3.3.3 Forecasted Wind Runs

Altering the wind direction changes the relative winds impeding or aiding the forward progress of the two aircraft. The averages (root mean square values) of the changes in the predicted relative wind magnitudes between the control run (Run 000) and the other runs are shown in Figure 3.3-11. The induced forecasted wind errors change the average relative wind by similar amounts for each of the aircraft. The largest increase is in Run 200.

In the control run both aircraft have a predicted 25 to 29 knot head wind during the first portion of their flights. In the Run 020, GHI1000 has a predicted 4 knot tail wind followed by a 10 knot head wind followed by a 15 knot head wind. QRS1000 has a 5 to 10 knot tail wind increasing to 20 knots at the end of the flight.

3.3.3.4 Forecasted Air Temperature Runs

As in the previous flight example, the forecasted air temperatures are unchanged in Runs 100, 200, 010, and 020. They are altered in Runs 001 and 002 only. As expected the wind magnitude RMS values in Figure 3.3-11 are zero for these runs. The winds are unaltered in the air temperature runs (Run 001 and Run 002).

3.3.3.5 Longitudinal Errors

The errors introduced into the forecasted winds and temperatures cause changes in the longitudinal errors in the aircraft trajectories as compared to the control run. The standard deviations of the changes are shown in Figure 3.3-12. This longitudinal error is increased the greatest in the Run 020. The longitudinal error is larger for QRS1000 than for GHI1000 in every run except Run 100, where their standard deviation is almost the same.

3.3.3.6 Conflict Predictions

For the GHI1000 and QRS1000 encounter, the URET notification sets are listed in Table 3.3-3 and Table 3.3-4. For this example, only red alerts were generated by URET, so only analyses AA and BA will be presented. Once again, the other two analyses, AB and BB, include both yellow and red alerts so would not be different for this example. As discussed previously, the aircraft only come within 14.09 nm horizontally; URET's red alerts predict conflicts (i.e. a violation of the separation minima or less than 5 nm) between these aircraft.

The control run, Run 000, has two notification sets. The first notification set predicts a conflict past the end of the track data for the two aircraft and is therefore discarded (see Section 2.2.3.1.1 for details on these rules). The second notification set predicts a conflict, but it is later retracted at 19:14:45 UTC due to a reconformance of a trajectory. All of the runs generate notification sets similar to the control run. Their times (notification start time, notification end time, predicted conflict start time, predicted conflict end time) are similar to each other. However, the second wind direction run, Run 020, generates two additional notification sets. This is consistent with the longitudinal error results in the previous section and presented in Figure 3.3-12, where Run 020 had the largest trajectory errors. These alerts are also retracted. As listed as the third notification set for Run 020 in Table 3.3-3, the first of these two additional notification sets is out of adherence. In the next table, Table 3.3-4, the adherence rule is ignored (Analysis BA), so Run 020 has two additional retracted false alerts instead of one.

3.3.3.7 Summary of Flight Example #3

The induced errors in wind forecast files have been shown to increase the relative wind magnitude error as measured between the control run and treatment runs. This was presented in Figure 3.3-11 and described in Section 3.3.3.3. The wind forecast errors translate to longitudinal trajectory errors for each of the flights as presented in Figure 3.3-12. These trajectory errors in turn cause URET to reconform its trajectories and induce additional retracted false alerts. For this example, the wind direction run, Run 020, is the most affected. It had the largest standard deviation of longitudinal error and two additional retracted false alerts, as shown in Table 3.3-4.

Flight Example #3 illustrates that the effects of the wind direction errors can increase the number of retracted false alerts. Similar to Flight Example #2 where the wind magnitude error caused the largest impact, the false alerts in this example are retracted when the URET reconforms the aircraft position to correct for longitudinal error in its trajectory predictions.

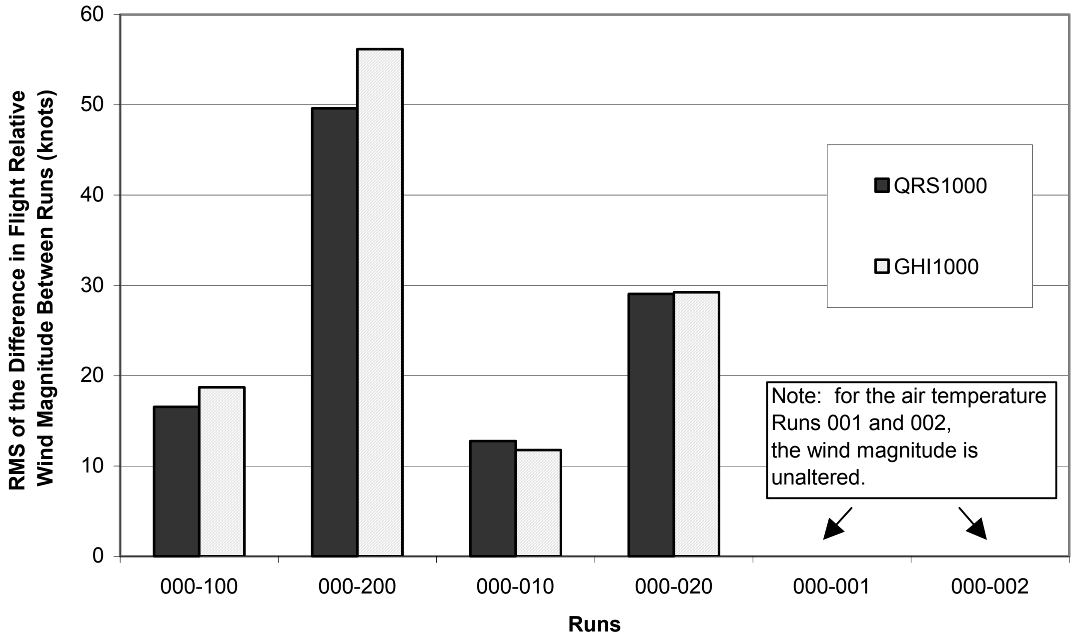


Figure 3.3-11 Example 3 RMS Values of the Differences in Relative Wind Magnitudes Between Runs

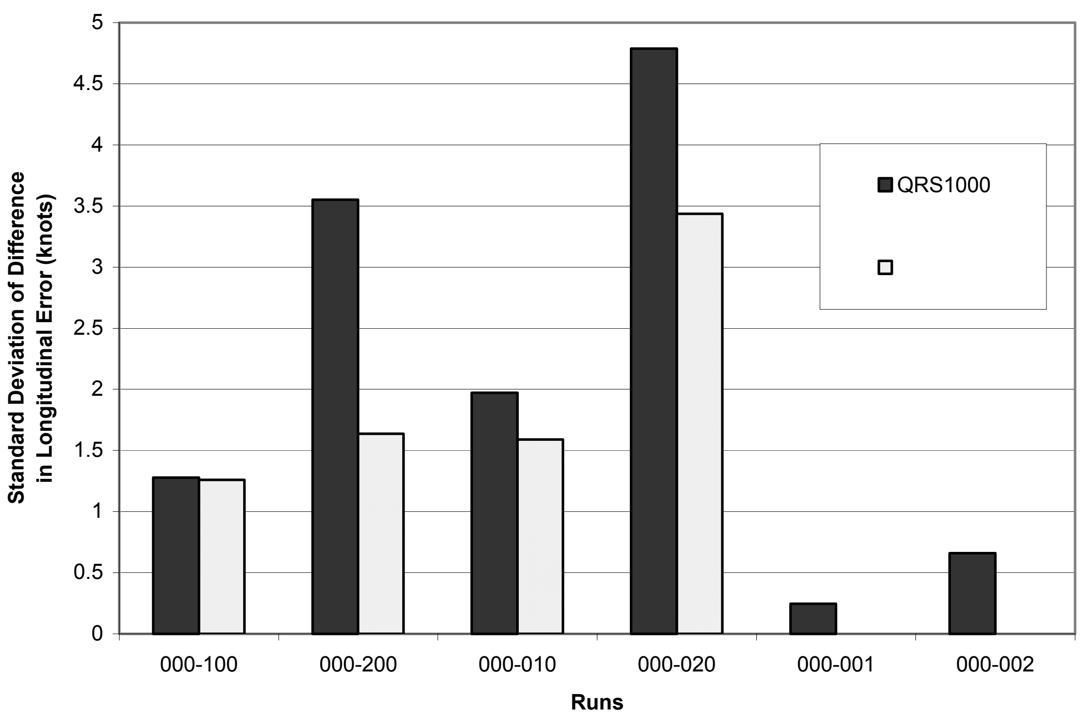


Figure 3.3-12 Example 3 Standard Deviations of the Differences in Longitudinal Errors Between Runs

Table 3.3-3 Example 3 Notification Sets for Encounter with Flights GHI1000 and QRS1000 – Adherence Rule Applied

Run and Notification Set Number	Set Start Time	Notification Set End Time	Predicted Conflict Start Time	Predicted Conflict End Time	Description
000AA-1	18:53:12	19:10:45	19:06:57	19:29:37	Discarded false alert due to no track data at PCST
000AA-2	19:10:45	19:14:45	19:22:16	19:45:06	Retracted false alert
100AA-1	18:52:44	19:10:45	19:06:29	19:29:08	Discarded false alert due to no track data at PCST
	19:10:45	19:10:51	19:20:12	19:46:10	Discarded false alert due to out of adherence
200AA-1	18:51:47	19:10:45	19:05:32	19:28:11	Discarded false alert due to no track data at PCST
200AA-2	19:12:21	19:15:03	19:25:30	19:47:58	Retracted false alert
010AA-1	18:53:38	19:10:45	19:07:23		Discarded false alert due to no track data at PCST
010AA-2	19:10:45	19:14:33	19:18:50	19:35:33	Retracted false alert
020AA-1	18:54:52	19:10:45	19:08:36	19:31:03	Discarded false alert due to no track data at PCST
020AA-2	19:10:45	19:12:21	19:14:29	19:22:10	Retracted false alert
020AA-3	19:12:21	19:14:45	19:19:42	19:39:09	Discarded false alert due to out of adherence
020AA-4	19:14:45	19:16:33	19:30:39	19:46:37	Retracted false alert
001AA-1	18:53:12	19:10:45	19:06:57	19:29:37	Discarded false alert due to no track data at PCST
001AA-2	19:10:45	19:14:45	19:22:16	19:45:06	Retracted false alert
002AA-1	18:53:24	19:10:45	19:07:09	19:29:37	Discarded false alert due to no track data at PCST
002AA-2	19:10:45	19:14:45	19:22:16	19:45:06	Retracted false alert

Table 3.3-4 Example 3 Notification Sets for Encounter with Flights GHI1000 and QRS1000 - Adherence Rule Ignored

Notification Set Number	Notification Set Start Time	Notification Set End Time	Predicted Conflict Start Time	Predicted Conflict End Time	Description
000BA-1	18:53:12	19:10:45	19:06:57	19:29:37	Discarded false alert due to no track data at PCST
000BA-2	19:10:45	19:14:45	19:22:16	19:45:06	Retracted false alert
100BA-1	18:52:44	19:10:45	19:06:29	19:29:08	Discarded false alert due to no track data at PCST
100BA-2	19:10:45		19:20:12	19:46:10	Retracted false alert
200BA-1	18:51:47	19:10:45	19:05:32	19:28:11	Discarded false alert due to no track data at PCST
200BA-2	19:12:21	19:15:03	19:25:30	19:47:58	Retracted false alert
010BA-1	18:53:38		19:07:23	19:29:51	Discarded false alert due to no track data at PCST
010BA-2	19:10:45	19:14:33	19:18:50	19:35:33	Retracted false alert
020BA-1	18:54:52	19:10:45	19:08:36	19:31:03	Discarded false alert due to no track data at PCST
020BA-2	19:10:45	19:12:21	19:14:29	19:22:10	Retracted false alert
020BA-3	19:12:21	19:14:45	19:19:42	19:39:09	Retracted false alert
020BA-4	19:14:45	19:16:33	19:30:39	19:46:37	Retracted false alert
001BA-1	18:53:12	19:10:45	19:06:57		Discarded false alert due to no track data at PCST
001BA-2	19:10:45	19:14:45	19:22:16	19:45:06	Retracted false alert
002BA-1	18:53:24	19:10:45	19:07:09	19:29:37	Discarded false alert due to no track data at PCST
002BA-2	19:10:45	19:14:45	19:22:16	19:45:06	Retracted false alert

3.3.4 Flight Example #4

Unlike the other examples that showed the impact on an encounter between aircraft, this example illustrates the URET effects of forecasted weather errors on an actual conflict in the scenario. Thus, both the missed and false alert conflict prediction accuracy will be examined.

3.3.4.1 Flight Paths

The JKL1000 flight is a Boeing 727 flying from Salt Lake City to Philadelphia at FL 330. The route is from Salt Lake City to a radial from the Bradford VORTAC in IL, to the Rosewood VORTAC in OH, onto J152 jetway, taking it to the Johnston VORTAC in PA, and from there using the BUNTS1 STAR into the Philadelphia airport. Midway in the route the aircraft is allowed to fly direct to the Johnston VORTAC.

The NOP1000 flight is a MD80 flying from St. Louis to Newark, NJ at FL 330. The route is a departure on the GATEWAY SID through the VORTACs at Rosewood, OH, Chardon, OH, Slate Run, PA, and finally on the WILLIAMSPORTS1 STAR into the Newark airport. In the test scenario, it starts in a climb at 21,200 feet. At 24,800 feet it is cleared to FL 290; at 28,400 feet the interim altitude is removed. It reaches level cruise at FL 330 at 19:01:40 and remains at this altitude until the end of the scenario.

3.3.4.2 Conflict

The two flights come into conflict at crossing courses at FL 330 at 19:16:50 UTC and remain in conflict for forty seconds. The minimum horizontal separation is 2.61 nm. Both flights are in cruise at FL 330 during the entire conflict. NOP1000 is flying a straight course, while JKL1000 turns and crosses the NOP1000 flight path, almost at right angles. This is reflected in the encounter angle of the conflict, which is 86 degrees. As the conflict ends, the NOP1000 flight passes behind JKL1000.

3.3.4.3 Forecasted Wind Runs

As described previously, errors in wind magnitude are introduced directly into the forecasted weather data for Runs 100 and 200 and indirectly by altering the wind directions in Runs 010 and 020. The differences from the control versus treatment runs in the relative wind magnitudes (i.e. wind magnitude projected onto the aircraft's route of flight) are shown in Figure 3.3-13. These errors are calculated individually for NOP1000 and JKL1000 for all its reported track positions and as in the previous flight examples reported as a RMS value. The largest RMS change (50 knots) is in Run 200 for the JKL1000 flight.

3.3.4.4 Forecasted Air Temperature Runs

As in the previous flight example, the forecasted air temperatures are unchanged in Runs 100, 200, 010, and 020. They are altered in Runs 001 and 002 only, yet as expected the wind magnitude RMS values in Figure 3.3-13 are zero for these runs. Thus, the winds are unaltered in the air temperature runs (Run 001 and Run 002).

3.3.4.5 Longitudinal Errors

The standard deviations of the changes in the longitudinal errors caused by the induced of weather errors are given in Figure 3.3-14. The errors are substantially larger for NOP1000 in Run 200 and Run 020, and somewhat less for the JKL1000 flight.

3.3.4.6 Conflict Predictions

The conflict predictions are listed in Table 3.3-5. The following subsections describe the results for the various runs.

3.3.4.6.1 Control Run

In the control run, Run 000, URET predicts one conflict between the two aircraft. A yellow alert is posted at 19:03:33 UTC (68613s) predicting a conflict starting at 19:16:18 UTC (69378s). At 19:16:33 UTC (69393s) the alert is changed to red and the conflict is predicted to start at 69393s (immediately). The actual conflict start time is 19:16:50 UTC (69410s). The yellow alert notification set has a warning time of 13min 17s and the red alert notification set has a warning time of 17s. In the AA and BA processing the alert is classified as LATE_MA; in the AB and BB processing the alert is classified as a STD_VA.

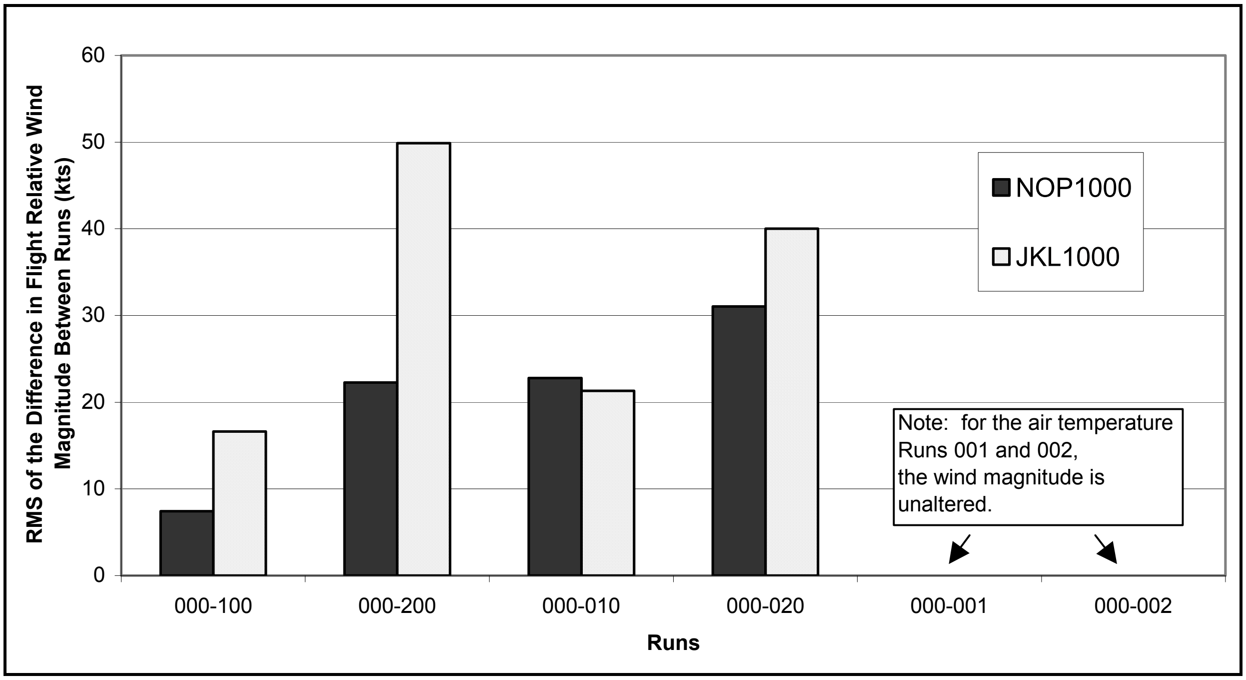


Figure 3.3-13 Example 4 RMS Values of the Differences in Relative Wind Magnitudes Between Runs

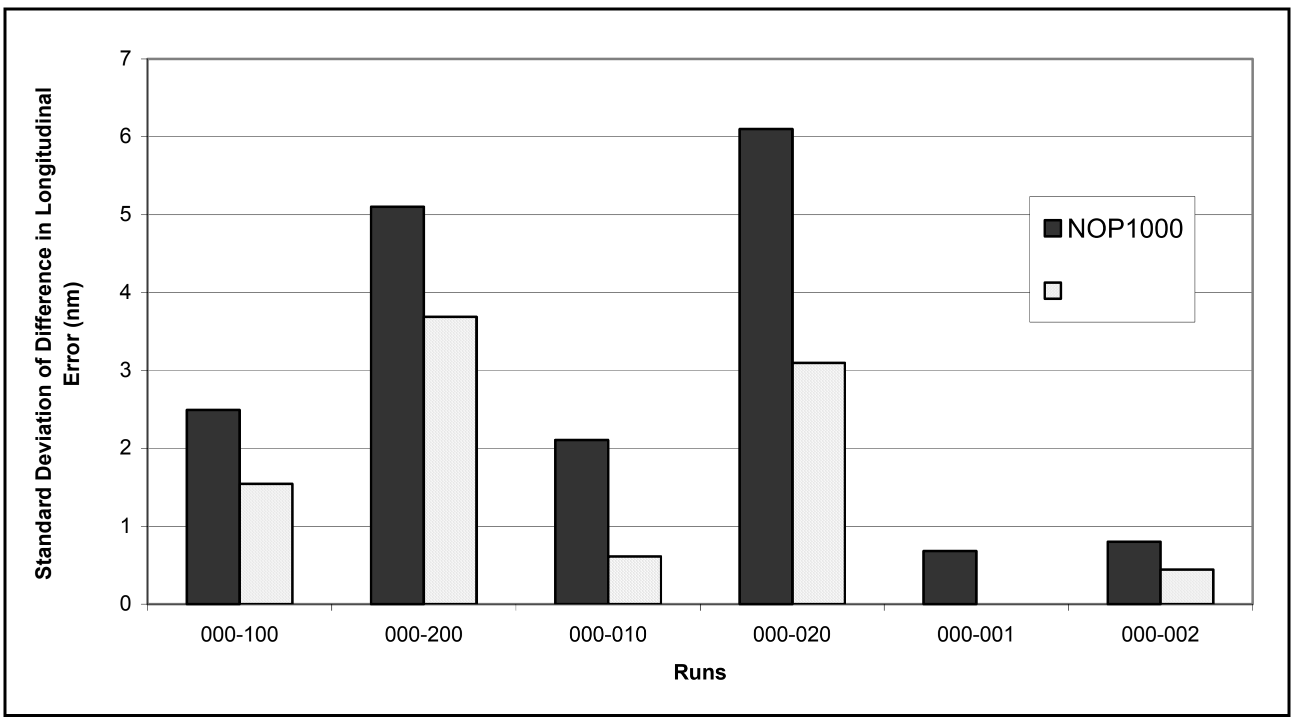


Figure 3.3-14 Example 4 Standard Deviations of the Differences in Longitudinal Errors between Runs

Table 3.3-5 Example 4 Notification Sets for Conflict with Flights JKL1000 and NOP1000

Run and Notification Set	Notification Set Start Time	Notification Set End Time	Predicted Conflict Start Time	Predicted Conflict End Time	Description
000AA-1	19:16:33	19:18:09	19:16:33	19:20:30	Late missed alert
100AA-1	19:03:33	19:04:12	19:15:27	19:20:54	Retracted false alert
100AA-2	19:16:33	19:18:10	19:16:33	19:20:31	Late missed alert
200AA-1	19:00:42	19:03:01	19:14:32	19:19:20	Retracted false alert
200AA-2	19:16:33	19:18:10	19:16:33	19:20:33	Late missed alert
010AA-1	19:03:33	19:03:57	19:15:54	19:21:22	Retracted false alert
010AA-2		19:18:09	19:16:33	19:20:31	Late missed alert
020AA-1	19:02:36	19:03:33	19:15:24	19:20:45	Retracted false alert
020AA-2	19:03:57	19:09:57	19:15:39	19:20:41	Retracted false alert
020AA-3	19:16:33	19:18:09	19:16:33	19:20:37	Late missed alert
001AA-1	19:03:33		19:16:13	19:21:25	Retracted false alert
001AA-2	19:16:33	19:18:09	19:16:33	19:20:31	Late missed alert
002AA-1	19:03:33	19:03:57	19:16:13	19:21:25	Retracted false alert
002AA-2	19:16:33	19:18:09	19:16:33	19:20:31	Late missed alert
000AB-1	19:03:33	19:18:20	19:16:18	19:21:22	Standard valid alert
100AB-1	19:02:15	19:04:12	19:15:27	19:19:40	Retracted false alert
100AB-2	19:06:57	19:10:56	19:16:18	19:20:46	Retracted false alert
100AB-3	19:16:21	19:18:20		19:21:18	Late missed alert
200AB-1	19:00:42	19:03:01	19:14:32	19:19:20	Retracted false alert
200AB-2	19:08:09	19:14:45	19:16:12	19:20:33	Retracted false alert
200AB-3	19:16:33	19:18:20	19:16:33	19:20:33	Late missed alert
010AB-1	19:03:23	19:07:00	19:16:37	19:19:36	Retracted false alert
010AB-2	19:16:33	19:18:20	19:16:33	19:20:31	Late missed alert
020AB-1	19:02:36	19:14:00	19:15:24	19:20:45	Retracted false alert
020AB-2	19:16:09	19:18:20	19:17:28	19:20:04	Late missed alert
001AB-1	19:03:33	19:14:00	19:16:13	19:21:25	Retracted false alert
001AB-2	19:16:33	19:18:20	19:16:33	19:20:31	Late missed alert
002AB-1	19:03:33	19:15:01	19:16:13	19:21:25	Retracted false alert
	19:16:33	19:18:20	19:16:33	19:20:31	Late missed alert
000BA-1	19:16:33	19:18:09	19:16:33	19:20:30	Late missed alert
100BA-1	19:03:33	19:04:12	19:15:27	19:20:54	Retracted false alert
100BA-2	19:16:33	19:18:10	19:16:33	19:20:31	Late missed alert
200BA-1	19:00:42	19:03:01	19:14:32	19:19:20	Retracted false alert
200BA-2	19:16:33	19:18:10	19:16:33	19:20:33	Late missed alert
010BA-1	19:03:33	19:03:57	19:15:54	19:21:22	Retracted false alert
010BA-2	19:16:33	19:18:09	19:16:33	19:20:31	Late missed alert
020BA-1	19:02:36	19:03:33	19:15:24	19:20:45	Retracted false alert
020BA-2	19:03:57	19:09:57	19:15:39	19:20:41	Retracted false alert
020BA-3	19:16:33	19:18:09	19:16:33	19:20:37	Late missed alert
001BA-1	19:03:33	19:03:57	19:16:13	19:21:25	Retracted false alert
001BA-2	19:16:33	19:18:09	19:16:33	19:20:31	Late missed alert
002BA-1	19:03:33	19:03:57	19:16:13	19:21:25	Retracted false alert
002BA-2	19:16:33	19:18:09	19:16:33	19:20:31	Late missed alert

Run and Notification Set	Notification Set Start Time	Notification Set End Time	Predicted Conflict Start Time	Predicted Conflict End Time	Description
000BB-1	19:03:33	19:18:20	19:16:18	19:21:22	Standard valid alert
100BB-1	19:02:15	19:04:12	19:15:27	19:19:40	Retracted false alert
100BB-2	19:06:57	19:10:56	19:16:18	19:20:46	Retracted false alert
100BB-3	19:16:21	19:18:20	19:16:21	19:21:18	Late missed alert
200BB-1	19:00:42	19:03:01	19:14:32	19:19:20	Retracted false alert
200BB-2	19:08:09	19:14:45	19:16:12	19:20:33	Retracted false alert
200BB-3	19:16:33	19:18:20	19:16:33	19:20:33	Late missed alert
010BB-1	19:03:23	19:07:00	19:16:37	19:19:36	Retracted false alert
010BB-2	19:16:33	19:18:20	19:16:33	19:20:31	Late missed alert
020BB-1	19:02:36	19:14:00	19:15:24	19:20:45	Retracted false alert
020BB-2	19:16:09	19:18:20	19:17:28	19:20:04	Late missed alert
001BB-1	19:03:33	19:14:00	19:16:13	19:21:25	Retracted false alert
001BB-2	19:16:33	19:18:20	19:16:33	19:20:31	Late missed alert
002BB-1	19:03:33	19:15:01	19:16:13	19:21:25	Retracted false alert
002BB-2	19:16:33	19:18:20	19:16:33	19:20:31	Late missed alert

The rules for processing these alerts are described in detail in Section 2.2.3.1.1. Briefly, for the Analysis AA and BA, red alerts are required for a notification set to be valid. The red alert is presented only 17s before the conflict starts. To be a valid alert, a warning time of five minutes is required. For the Analysis AB and BB, both red and yellow alerts can be used. In this example, the yellow alert notification set is posted well beyond the five-minute warning time requirement.

3.3.4.6.2 Forecasted Wind Runs

For the four runs with induced wind error (i.e. Run 100, 200, 010, and 020), all post an alert for the conflict, but in every case it is too late to satisfy the warning time requirement of five minutes. They also all post timely red alerts, but all are retracted due to reconformed trajectories.

3.3.4.6.3 Forecasted Air Temperature Runs

The two runs with induced air temperature error post an alert but withdraw it before the conflict starts, resulting in a retracted false alert. They later post an alert for the conflict, but it is also too late to satisfy the warning time requirement of five minutes.

3.3.4.7 Summary of Flight Example #4

The experiment induced both wind and air temperature errors to the forecasted winds URET uses to build aircraft trajectories and make conflict predictions. For this example, the two aircraft designated as NOP1000 and JKL1000 have a test conflict, which cross paths and violate separation standards at almost 90 degrees with a minimum horizontal separation of 2.6 nm. All the weather treatment runs miss the conflict by presenting an alert too late (less than five minutes from the start of the conflict). The control run presents a yellow alert with about 13 minutes of warning time, forming a standard valid alert under Analysis AB and BB. For this example, the treatment runs also generate several retracted false alerts. This example illustrates that the induced weather forecast errors can generate late missed alerts as well as retracted false alerts. However, we know from the overall results in Section 3.2 the impact on missed alert probability is not significant for the entire scenario of conflicts and treatment runs.

3.3.5 Summary of Flight Example Results

The flight examples have illustrated the effects of the errors introduced into the predicted wind magnitudes, the predicted wind directions, and the predicted air temperatures. For Flight Example #1, the effects are listed in order of occurrence: the errors in predicted wind magnitudes lead to errors in predicted relative winds, which lead to errors in predicted aircraft ground speeds. These ground speed errors lead to errors in the four dimensional trajectories, mainly longitudinally. Similarly errors in predicted wind directions lead to errors in predicted relative winds and so forth. The errors in predicted air temperatures lead to errors in the predicted rates of climb, which lead to errors in the trajectories.

For Flight Example #1, the errors in the trajectories are illustrated in the trajectories and track positions in Figure 3.3-5 to Figure 3.3-8. The wind errors increase the longitudinal errors but do not affect the lateral errors. The vertical error is increased as well but only very slightly. The conflicts predicted (or not predicted) are based on the trajectories. Trajectory errors cause conflict prediction errors. The ground speed errors in the trajectories cause them to deviate from the radar track. In turn, the differences between the track and the predicted positions cause URET to reconform and calculate a new trajectory. Conflicts predicted in error on the basis of the old trajectory are deleted (retracted) when the old trajectory is replaced by the new trajectory after a reconformance. The errors introduced into the weather data increase the number of reconformances and the number of false alerts. Most of the false alerts are retracted upon reconformance to the radar track. Similar results are found in the Flight Examples #2 and #3, where retracted false alerts are generated for both the wind magnitude and direction runs.

The Flight Example #4, illustrating a conflict, shows that the reconformance, induced by the weather errors, can also cause the retraction of a valid alert and thus a missed alert error.

4 Conclusion

4.1 Overview of the Experiment

The objective of this study was to evaluate what impact weather forecast errors have on URET trajectory and conflict predictions. As described in detail in Section 2, wind and air temperature errors were induced by altering URET's weather forecast files. A comprehensive analysis followed in Section 3.

In summary, the experiment used about two hours of traffic data recorded at the ZID ARTCC in May 1999. The flights were time shifted to generate a sufficient number of test conflicts using a genetic algorithm technique developed by CPAT (see Section 2.1.3.1.2 for details). This time-shifted scenario was used as input to the URET Prototype. To induce weather forecast error, the weather input file (RUC) was altered by adding 20 or 60 knots to the wind magnitude, 45 or 90 degrees to the wind direction, and 5 or 15 degrees Kelvin to the air temperature (see Section 2.1.2). This produced seven URET runs for the experiment – the unaltered control run and six treatment runs (see Table 2.1-1 for listing of runs). The analysis consisted of comparing the treatment runs against this control run and is presented in Section 3.

4.2 Statistical Analysis Conclusions

URET's trajectory predictions were analyzed for statistically significant effects. For both wind magnitude levels (20 and 60 knots), horizontal trajectory error and its along path component, longitudinal trajectory error, were statistically significant for all look-ahead times (i.e. 0 to 20 minutes) and for both level and in-transition phase-of-flight. Similar results occurred for the wind direction runs, except for the in-transition Run 010 (45 degree). The air temperature runs did not differ statistically from the control run. As illustrated in Table 3.1-8, the errors in trajectory predictions cause URET to produce more trajectories per flight because it reconforms to correct for the longitudinal error. This is consistent with the trajectory error results, since it is only demonstrated in the wind treatment runs.

Similarly, the missed and false alert errors were evaluated for each run and then comparisons were performed. The complete comparison results are presented in the tables in Section 3.2.2. There was no evidence that the probability of missed alert differed between the control and any of the treatment runs. However, there was a difference detected in the probability of false alert error in some of the treatment runs.

The air temperature treatment runs had no evidence of a difference for either missed or false alert error. For the wind magnitude and direction runs, the false alert probability was statistically different, but the differences were not very high and ranged from 0.01 to 0.06. Furthermore, the difference was dominated by the number of retracted false alerts. This was illustrated in Figure 3.2-1, Figure 3.2-2, Figure 3.2-3, and Figure 3.2-4 that plotted the percentage in false alert differences for both standard and retracted false alerts for all for analyses. In addition, an analysis was performed comparing the warning times on the valid alerts common between runs. This analysis was repeated for the predicted conflict start times. In all cases, these comparisons did not show a significant difference between the control and treatment runs.

4.3 Flight Analysis Conclusions

The statistical analysis provided evidence that the wind magnitude and direction levels had a modest impact on retracted false alerts. Four flights and their encounters with other aircraft were analyzed to help determine the causes of this overall effect. In summary, the error added to the forecasted wind data causes additional errors in predicted positions. The new errors are principally along path or longitudinal errors caused by inaccurate predictions of ground speed. The increase in longitudinal error is consistent with the trajectory accuracy results. Vertical position errors are caused primarily by errors in predicted climb rate. The predicted climb rates are affected only slightly by errors in predicted winds, while the predicted climb angles are affected somewhat more. This small vertical effect is not statistically significant as shown above and in Section 3.1. The position errors cause URET to rebuild trajectories more frequently resulting in retractions in conflict predictions. Thus, the number of retracted false alerts are increased. The Flight Example #4 (see Section 3.3.4) demonstrates that an individual flight may be greatly impacted, but the aggregate effect on missed alert probability was not statistically significant. The URET trajectory reconformance logic correctly adjusted its trajectories to avoid missing conflict predictions. As expected, this same reconformance logic caused more retracted false alerts to be generated.

4.4 Operational Recommendations

Operationally, weather forecasts may be inaccurate due to the presence of highly dynamic weather or outages in the interfaces to the NWS. This study showed that induced errors, as high as 60 knots in wind magnitude and 90 degrees in wind direction, had a modest effect on URET predictions. Therefore, a controller suspecting errors in the input wind forecast should expect only a modest impact on URET predictions. The impact would mainly be a moderate increase in the number of retracted false alerts, yet no overall affect on missed alert error. This is consistent with [Lindsay, 1997a], which reported URET predictions still have utility under degraded weather forecast errors. If a controller notices an increase in retractions, it may be symptomatic of inaccurate wind forecasts, which should be investigated.

4.5 Future Research

Although the effect of weather errors on URET Prototype's predictions was shown to be minor, future research should confirm the applicability of these results to the production version of URET (Core Capability Limited Deployment, CCLD). Other researchers have already thoroughly studied the errors themselves in the weather forecast files, such as [Cole et al., 2000] and [Sherry, 1999], but another complementary study should focus on the disparity of the weather forecast files as they change from hour to hour and their forecast age as input to URET. In addition, future research should investigate the impact of convective weather on URET predictions.

5 List of Acronyms

ACARS	Aircraft Communication Addressing and Reporting System
ACB-330	Simulation and Analysis Group
AEC	Algorithmic Evaluation Capability
ANOVA	Analysis of Variance
ARTCC	Air Route Traffic Control Center
ASC	Ascending
ASCII	American Standard Code for Information Exchange
ATM	Air Traffic Management
AWIPS	Advanced Weather Interactive Process System
CAASD	Center for Advanced Aviation System Development
CCLD	Core Capability Limited Deployment
CONUS	Continental United States
CP	Conflict Probe
CPAT	Conflict Probe Assessment Team
CTAS	Center TRACON Automation System
DSC	Descending
DST	Decision Support Tool
FAA	Federal Aviation Administration
FL	Flight Level
FFTIL	Free Flight Technology Integration Laboratory
GA	Genetic algorithm
GDS	Grid Description Section
gcc	GNU C/C++ Compiler
GNU	GNU's Not Unix
gpm	geopotential meters
GRIB	Gridded Binary
HCS	Host Computer System
JMP	SAS Statistical Package
K	Degrees Kelvin
km	Kilometers
LEV	Level Flight
LFT	Left
libg	GNU C/C++ Libraries
LMATM	Lockheed Martin Air Traffic Management
m	Meters
min	Minutes
Mb	Millibars
MDCRS	Meteorological Data Collection and Reporting System
Mhz	Megahertz
M/s	Meters per second
NAS	National Airspace System
NET	Notification End Time
Nm	Nautical miles
NST	Notification Start Time
NWS	National Weather Service
ORS	Onboard Route Segments

PCST	Predicted Conflict Start Time
PCET	Predicted Conflict End Time
PDS	Product Description Section
PoF	Phase-of-Flight
RHT	Right
RMS	Root Mean Square
RUC	Rapid Update Cycle
S	Seconds
SAS	Statistical Analysis System
SID	Standard Instrument Departure
SSG	State Segment
STAR	Standard Terminal Arrival Route
STR	Straight
STR-LEV	Straight and Level Flight
STR-TRAN	Straight and In-Transition flight
TRACON	Terminal Radar Approach Control
TRAN	In Transition flight
TRN-LEV	Turn and Level flight
TRN-TRAN	Turn and In-Transition flight
URET	User Request Evaluation Tool
UTC	Universal Coordinated Time
VHF	Very High Frequency
VORTAC	VHF Omni-Directional Range / Tactical Air Navigation
WARES	Winds Aloft Requirements Evaluation System
WJHTC	William J. Hughes Technical Center
ZAU	Chicago Air Route Traffic Control Center
ZID	Indianapolis Air Route Traffic Control Center

6 References

[Bashioum and Mayo, 1997] Bashioum, Douglas L., Mayo, John J. III (October 1997), "User Request Evaluation Tool (URET) D3 Trajectory Modeling Algorithmic Definition, MITRE Technical Report," MITRE/CAASD Technical Report, MTR97W0000072R1.

[Bilimoria , 2001] Bilimoria, Karl, (May-June 2001), "A Methodology for the Performance Evaluation of a Conflict Probe," *Journal of Guidance, Control, and Dynamics*, Vol. 24 (3).

[Bowen, 1997] Bowen, Karen C. (February 1997), "User Request Evaluation Tool with Display System Replacement Computer Human Interface Requirements," MITRE/CAASD Technical Report, MTR 97W0000014.

[Brudnicki , 1998] Brudnicki, D., Arthur, W., Lindsay, K. (April 1998), "URET Scenario-based Functional Performance Requirements Document," MITRE/CAASD Technical Report, MTR98W0000044.

[Cale et al., 1998] Cale, M., Paglione, M., Ryan, Dr. H., Timoteo, D., Oaks, R. (April 1998), *URET Conflict Prediction Accuracy Report*, (DOT/FAA/CT-TN98/8), WJHTC/ACT-250.

[Cale et al., 2001] Cale, M., Liu, S., Oaks, B., Paglione, M., Ryan, Dr. H., Summerill, S. (December 3-7, 2001), "A Generic Sampling Technique for Measuring Aircraft Trajectory Prediction Accuracy", *Presented at the 4th USA/Europe Air Traffic Management R&D Seminar*, Santa Fe, NM.

[Cole et al., 2000] Cole, Rodney E., Green, Steve, Jardin, Matt, Schwartz, Barry E., Benjamin, Stanley G., (June 13-16, 2000), "Wind Prediction Accuracy for Air Traffic Management Decision Support Tools," *Presented at the 3rd USA/Europe Air Traffic Management R&D Seminar*, Napoli, Italy.

[Devore, 2000] Devore, Jay L. (2000), *Probability and Statistics for Engineering and the Sciences, Fifth Edition*.

[Heim, 1999] Heim, J. J. (November 29, 1999), "URET Rapid Update Cycle Data/URET Interface Description," MITRE Memo No. F042-M-037 to J. J. Mayo and M. E. Tucker.

[Hollander and Wolfe, 1999] Hollander, M., Wolfe, D. A. (1999), "*Nonparametric Statistical Methods", Second Edition*, John Wiley & Sons, Inc.

[Jaarsma et al., 1997] Jaarsma, Jan, Jacobs, Scott, McFarland, Dr. Alvin, (1999), "User Request Evaluation Tool (URET) Delivery 3.1 System Test Tools," MITRE/CAASD Technical Report, WN99W0000010.

[Johnson and Dey, 2000] Johnson, Joseph E., Dey, Clifford H. (December 15, 2000), *Office Note 388 GRIB (Edition 1) The WMO Format for the Storage of Weather Product Information and the Exchange of Weather Product Messages in Gridded Binary Form as Used by NCEP Central Operations*, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, National Centers for Environmental Prediction. (Available at: ftp://ftp.ncep.noaa.gov/pub/nws/nmc/docs/gribedl)

[Kachigan, 1986] Kachigan, Sam Kash (1986), *Statistical Analysis: An Interdisciplinary Introduction to Univariate & Multivariate Methods*.

[Kelly, 2000] Kelly, Betty, Paglione, Mike, Ryan, Dr. Hollis (November 2000), "URET CCLD Accuracy Scenario Weather Forecast Deviation Study."

[Lindsay, 1997a] Lindsay, Kenneth S. (April 21, 1997), "Sensitivity of AEC/URET Performance to Wind Data", MITRE/CAASD Technical Memorandum, F042-M-087.

[Lindsay, 1997b] Lindsay, Kenneth S. (June 1997), "Xeval Users Manual", MITRE/CAASD Technical Report, MTR 97W0000030.

[Lindsay, 1998] Lindsay, Kenneth S. (June 1998), "Algorithmic Evaluation Process and Tools for the User Request Evaluation Tool", MITRE/CAASD Technical Report, MTR 98W0000029.

[LMATM, 2001] Lockheed Martin Air Traffic Management, (June 1, 2001), "User Request Evaluation Tool Core Capability Limited Deployment Software Requirements Specification Conflict Probe Algorithmic Data Dictionary, Volume V," Lockheed Martin Corporation, Rockville, MD.

[LMATM, 2002] Lockheed Martin Air Traffic Management, (April 25, 2002), "User Request Evaluation Tool (URET) Core Capability Limited Deployment (CCLD) Test Report for the Formal Accuracy Test, Volume III," Lockheed Martin Corporation, Rockville, MD.

[Montgomery, 1997] Montgomery, Douglas C. (1997), *Design and Analysis of Experiments, Fourth Edition*.

[Oaks and Paglione, 2001] Oaks, R., Paglione, M. (Fall 2001), "Generation of Realistic Air Traffic Scenarios Based on Recorded Field Data," *46th Annual Air Traffic Control Association Conference Proceedings*, Arlington, VA, pp.142-146.

[Oaks, 2002] Oaks, Robert D. (August 2002), "A Study on the Feasibility of Using a Genetic Algorithm to Generate Realistic Air Traffic Scenarios Based on Recorded Field Data," presented at the AIAA Guidance, Navigation, and Control Conference, Monterey, CA.

[Oaks and Paglione, 2002] Oaks, Robert D., Paglione, Mike (October 2002), "Generation of Realistic Air Traffic Scenarios Using a Genetic Algorithm," to be presented at the 21st Digital Avionics Systems Conference, Irvine, CA.

[Paglione et al., 1997] Paglione, M., Ryan, Dr. H., Kazunas, S., Cale, M. (Fall 1997), "Generic Metrics for the Conflict Probe Tools Developed for Free Flight", *42nd Annual Air Traffic Control Association Conference Proceedings*, Washington, DC.

[Paglione et al., 1999a] Paglione et al. (May 1999), "Trajectory Prediction Accuracy Report: URET/CTAS", (DOT/FAA/CT-TN99/10), WJHTC/ACT-250.

[Paglione et al., 1999b] Paglione, M., Cale, M., Ryan, Dr. H. (1999), "Generic Metrics for the Estimation of the Prediction Accuracy of Aircraft to Aircraft Conflicts by a Strategic Conflict Probe Tool", *Air Traffic Quarterly, Air Traffic Control Association*, Volume 7(3) 147-165.

[Paglione, 2000] Paglione, M. (July 24, 2000), "Quick Example of Benefits of Pairing Trajectory Accuracy Data,"Briefing at Performance Technical Interchange Meeting at Lockheed Martin Air Traffic Management.

[Paglione et al., 2000] Paglione, M., R. D. Oaks, H. F. Ryan, J. S. Summerill, January, 27, 2000, "Description of Accuracy Scenarios for the Acceptance Testing of the User Request Evaluation Tool (URET) / Core Capability Limited Deployment (CCLD), Final," WJHTC/ACT-250.

[Paglione and Summerill, 2000] Paglione, M. M., J. S. Summerill, (August 23, 2000), "Study of Lateral Adherence Thresholds for User Request Evaluation Tool Core Capability Limited Deployment Accuracy Acceptance Testing," WJHTC/ACT-250.

[Paglione et al., 2001] Paglione et al., (November 2001), "Measurement of Aircraft Trajectory Prediction Accuracy of Air Traffic Decision Support Tools," *46th Annual Air Traffic Control Conference Proceedings*, Washington, D.C.

[Paglione et al., 2002] Paglione, M., Ryan, H. R., PhD., Liu, S., "Accuracy Metrics for Conflict Probe Testing," Internal Memorandum, WJHTC/ACB-330.

[Sawilowsky, 1990] Sawilowsky, S. S. (1990), *Nonparametric Tests of Interaction in Experimental design, Review of Educational research*, Spring 1990, Vol. 60, No. 1, pp. 91-126.

[Schwartz and Benjamin, 1998] Schwartz, Barry E., Benjamin, Stanley G. (September 1998), "A Study of the Accuracy of RUC-1 and RUC-2 Wind and Aircraft Trajectory Forecasts Using ACARS Observations," NOAA Technical Memorandum ERL FSL-22.

[Sherry, 1999] Sherry, Joseph E. (July 1999), "Rapid Update Cycle (RUC) Winds Aloft Forecast Accuracy and Recommended ATM Requirements (Revision 1)," MITRE/CAASD Technical Report, MTR 98W0000098R1.

[Wanke, 1997] Wanke, Craig (June 17-20, 1997), "Using Air-Ground Data Link to Improve Air Traffic Management Decision Support System Performance," MITRE Corporation, *Presented at the 1st USA/Europe Air Traffic Management R&D Seminar*, Saclay, France.

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

7 Index

A

actual conflict start time (ACST), 33, 35, 44, 90
adherence, xvi, 21, 31, 33, 34, 35, 36, 37, 39, 40, 55, 56, 57, 58, 59, 60, 63, 65, 68, 70, 72, 80, 82, 84, 86, 88, 89, 101
Advanced Interactive Process System (AWIPS), 13, 97
air temperature, v, ix, xii, xiii, xv, 1, 5, 6, 13, 14, 17, 20, 24, 25, 51, 52, 54, 55, 61, 63, 66, 69, 72, 80, 82, 85, 90, 93, 94, 95
Air Traffic Database, xv, 7, 8
Air traffic scenario, v, ix, xi, xiv, 3, 5, 6, 7, 8, 9, 22, 23, 25, 32, 34, 36, 41, 73, 77, 81, 82, 83, 85, 89, 90, 93, 95, 99, 100
aircraft-to-aircraft, 5, 7, 8, 9, 22, 31

B

binomial distribution, 41

C

Cat, 9, 10
chi-squared distribution, 42
conflict, v, ix, xii, xiii, xiv, xv, xvi, 1, 2, 3, 4, 5, 7, 8, 9, 10, 12, 21, 22, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 43, 44, 55, 56, 57, 58, 59, 60, 61, 63, 65, 66, 68, 70, 71, 72, 73, 80, 81, 82, 84, 86, 88, 89, 90, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101
conflict prediction, v, ix, 1, 2, 3, 4, 5, 10, 22, 31, 32, 33, 35, 36, 43, 55, 56, 57, 58, 59, 60, 63, 65, 68, 70, 72, 73, 80, 81, 82, 89, 90, 93, 94, 95, 96
Conflict Probe (CP), v, ix, 1, 32, 34, 97, 99, 100, 101
Conflict Probe Assessment Team (CPAT), v, ix, xi, 1, 3, 4, 5, 7, 9, 10, 19, 21, 22, 32, 43, 95, 97
conflict start time, 33, 44, 71, 72, 82
Continental United States (CONUS), 13, 14, 97
control run, v, ix, 1, 5, 23, 26, 29, 31, 36, 43, 44, 47, 49, 51, 53, 54, 56, 57, 58, 59, 60, 64, 66, 69, 70, 71, 72, 73, 77, 78, 80, 82, 85, 86, 90, 93, 95

D

discarded false alert, 41, 84, 88, 89

E

encounter, ix, xiii, xvi, 8, 22, 31, 33, 37, 38, 43, 56, 57, 58, 59, 72, 80, 81, 82, 83, 84, 85, 86, 88, 89, 90, 96

F

false alert, v, ix, 31, 32, 33, 34, 35, 37, 38, 40, 41, 42, 43, 55, 56, 57, 58, 59, 60, 61, 63, 65, 66, 68, 70, 72, 80, 81, 83, 84, 86, 88, 89, 92, 93, 94, 95, 96
Free Flight Integration Laboratory (FFTIL), 21, 97

G

Gaussian distribution, 23, 25
genetic algorithm (GA), v, ix, 95, 97
getRUC, xi, 19
gribw, xi, 19
Gridded Binary (GRIB), 13, 19, 97, 99

H

horizontal error, 23, 26, 49, 50, 52, 53, 54, 55
Host Computer System (HCS), v, 6, 21, 22, 30, 31, 32, 33, 34, 82, 97

I

Indianapolis ARTCC (ZID), v, ix, 5, 7, 73, 95, 98

L

late missed alert, 33, 72, 92, 93
lateral error, 29, 31, 49, 51, 53, 54, 55, 81, 94
Lockheed Martin Air Traffic Management (LMATM), 9, 21, 32, 33, 36, 97, 100, 101
longitudinal error, ix, 23, 49, 51, 53, 54, 55, 77, 81, 82, 86, 90, 94, 95, 96
look-ahead time, 3, 4, 22, 25, 26, 28, 47, 49, 50, 51, 52, 53, 55, 95

M

missed alert, ix, 32, 33, 34, 37, 38, 39, 40, 41, 42, 43, 56, 57, 58, 59, 60, 63, 64, 65, 66, 68, 69, 72, 93, 94, 95, 96
MITRE CAASD, vii, ix, 1, 3, 4, 5, 6, 9, 21, 97, 99, 100, 101
modRUC, xi, 20
muted alert, 22, 32

N

nonparametric statistics, 23, 25, 26, 45
normal distribution, 23, 25
notification end time (NET), 36, 86, 97

O

Oracle, 7, 9, 19

P

parametric statistics, 1, 23
phase-of-flight (PoF), 47, 49, 50, 51, 53, 54, 55, 98
predicted conflict end time (PCET), 86, 98
predicted conflict start time (PCST), 3, 33, 34, 35, 36, 44, 70, 71, 72, 82, 86, 88, 89, 95, 98

Q

quantile-quantile plot (Q-Q), xiv, 23, 24

R

Rapid Update Cycle (RUC), v, ix, xi, xv, 3, 4, 5, 6, 13, 14, 15, 16, 17, 18, 19, 20, 23, 24, 25, 47, 49, 51, 54, 95, 98, 99, 101
red alert, 22, 31, 32, 55, 56, 57, 58, 59, 60, 63, 65, 68, 72, 80, 86, 90, 93
retracted false alert, v, ix, 33, 56, 58, 59, 60, 61, 64, 66, 69, 72, 81, 83, 86, 93, 94, 95, 96
Run 000, xiv, xv, 4, 6, 9, 14, 23, 27, 29, 30, 31, 43, 44, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 66, 67, 68, 69, 71, 72, 73, 80, 81, 82, 85, 86, 90
Run 001, 6, 51, 52, 53, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 66, 67, 68, 69, 82, 85, 90
Run 002, 6, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 66, 67, 68, 69, 70, 82, 85, 90
Run 010, 6, 23, 49, 50, 51, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 71, 73, 82, 85, 90, 93, 95
Run 020, 6, 49, 50, 51, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 82, 85, 86, 90, 93
Run 100, xv, 6, 14, 16, 17, 18, 23, 29, 30, 31, 47, 48, 49, 54, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 73, 77, 81, 82, 85, 86, 90, 93
Run 200, v, vii, xiv, 1, 6, 16, 17, 18, 23, 27, 43, 44, 47, 48, 49, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 73, 74, 77, 78, 81, 82, 83, 85, 90, 93

S

Statistical Analysis System (SAS), xi, xiv, 23, 26, 27, 29, 31, 47, 49, 97, 98

T

track data, xiv, 3, 4, 6, 7, 8, 22, 34, 36, 53, 73, 74, 75, 76, 77, 78, 81, 82, 86, 88, 89, 90, 94
trajectory data, v, ix, xi, xii, xiv, xv, 1, 2, 3, 4, 5, 6, 14, 21, 22, 23, 24, 25, 28, 29, 45, 47, 49, 51, 53, 54, 55, 73, 77, 78, 83, 86, 94, 95, 96, 99, 100, 101
treatment run, v, ix, 1, 5, 6, 23, 26, 29, 31, 36, 43, 44, 53, 54, 60, 63, 65, 66, 68, 69, 70, 71, 72, 73, 82, 86, 90, 93, 95

U

univariate statistics, xi, 23, 26, 47, 49, 100
User Request Evaluation Tool (URET), v, vii, ix, x, xv, 1, 3, 4, 5, 6, 9, 10, 13, 14, 21, 22, 24, 31, 32, 33, 34, 35, 36, 54, 55, 72, 73, 77, 78, 80, 81, 82, 85, 86, 89, 90, 93, 94, 95, 96, 98, 99, 100, 101

V

valid alert, 31, 32, 33, 34, 37, 38, 39, 41, 42, 43, 44, 55, 57, 59, 60, 70, 71, 72, 92, 93, 94, 95

W

warning time, 3, 33, 43, 44, 45, 55, 70, 72, 90, 93, 95
wgrib, xi, 19, 20
Wilcoxon signed-rank test, 23, 25, 26, 29, 30, 31, 45, 47, 49, 53, 71
wind direction, v, ix, 1, 5, 13, 20, 24, 25, 49, 54, 55, 72, 74, 75, 82, 85, 86, 94, 95, 96
wind magnitude, v, ix, 1, 5, 13, 20, 23, 24, 25, 30, 47, 49, 54, 72, 73, 75, 77, 80, 81, 82, 83, 85, 86, 90, 94, 95, 96

Y

yellow alert, 22, 31, 32, 55, 58, 59, 65, 68, 72, 80, 82, 83, 86, 90, 93

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

te technical note techn

User Request Evaluation Tool (URET) Conflict Probe Sensitivity to Weather Forecast Errors

APPENDIX A: Box Plots
APPENDIX B: Median Plots
APPENDIX C: Plots of Preliminary Data
APPENDIX D: Detailed Flight Example

Mike M. Paglione
Lori Charles
Shurong Liu
Robert D. Oaks
Hollis F. Ryan, Ph.D.
J. Scott Summerill

October 2002

DOT/FAA/CT-TN02/tbs

U.S. Department of Transportation
Federal Aviation Administration

William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

Table of Contents

A. Appendix A - Box Plots A-1
A.1.1 JMP Box Plots for Level Flight A-1
A.1.2 JMP Box Plots for In-Transition Flight A-10
B. Appendix B - Median Plots B-1
B.1.1 Plots for Level Phase of flight B-1
B.1.2 Median Plots for In-Transition Phase of Flight B-4
C. Appendix C - Preliminary Wind Data Analysis C-1
D. Appendix D – Detailed Flight Example D-1

List of Figures

Figure A1- 1 Quantile Plots of Error Measurement for Run 000-100 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-1
Figure A1- 2 Quantile Plots of Error Measurement for Run 000-200 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-2
Figure A1- 3 Quantile Plots of Error Measurement for Run 100-200 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-3
Figure A1- 4 Quantile Plots of Error Measurement for Run 000-010 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-4
Figure A1- 5 Quantile Plots of Error Measurement for Run 000-020 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-5
Figure A1- 6 Quantile Plots of Error Measurement for Run 010-020 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-6
Figure A1- 7 Quantile Plots of Trajectory Error Measurement for Run 000-001 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-7
Figure A1- 8 Quantile Plots of Trajectory Error Measurement for Run 000-002 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-8
Figure A1- 9 Quantile Plot of Error Measurement for Run 001-002 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-9
Figure A1- 10 Quantile Plots of Error Measurement for Run 000-100 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-10
Figure A1- 11 Quantile Plots of Error Measurement for Run 000-200 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-11
Figure A1- 12 Quantile Plots of Error Measurement for Run 100-200 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-12
Figure A1- 13 Quantile Plots of Error Measurement for Run 000-010 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-13
Figure A1- 14 Quantile Plots of Trajectory Error Measurement for Run 000-020 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-14
Figure A1- 15 Quantile Plots of Trajectory Error Measurement for Run 010-020 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-15
Figure A1- 16 Quantile Plots of Error Measurement for Run 000-001 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-16
Figure A1- 17 Quantile Plots of Error Measurement for Run 000-002 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-17
Figure A1- 18 Quantile Plots of Error Measurement for Run 001-002 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error) A-18
Figure B1- 1 Median Plots for Wind Magnitude (Horizontal, Vertical, Longitudinal, Lateral Error) B-1
Figure B1- 2 Median Plots for Wind Direction (Horizontal, Vertical, Longitudinal, Lateral Error) B-2
Figure B1- 3 Median Plots for Air Temperature (Horizontal, Vertical, Longitudinal, Lateral Error) B-3
Figure B1- 4 Median Plots for Wind Magnitude (Horizontal, Vertical, Longitudinal, Lateral Error) B-4
Figure B1- 5 Median Plots for Wind Direction (Horizontal, Vertical, Longitudinal, Lateral Errors) B-5
Figure B1- 6 Median Plots for Air Temperature (Horizontal, Vertical, Longitudinal, Lateral Error) B-6
Figure C1- 1 Box Plot of Wind Magnitude Data by Pressure Level for ZID Center (05/26/1999) C-1

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

A. Appendix A - Box Plots

A.1.1 JMP Box Plots for Level Flight

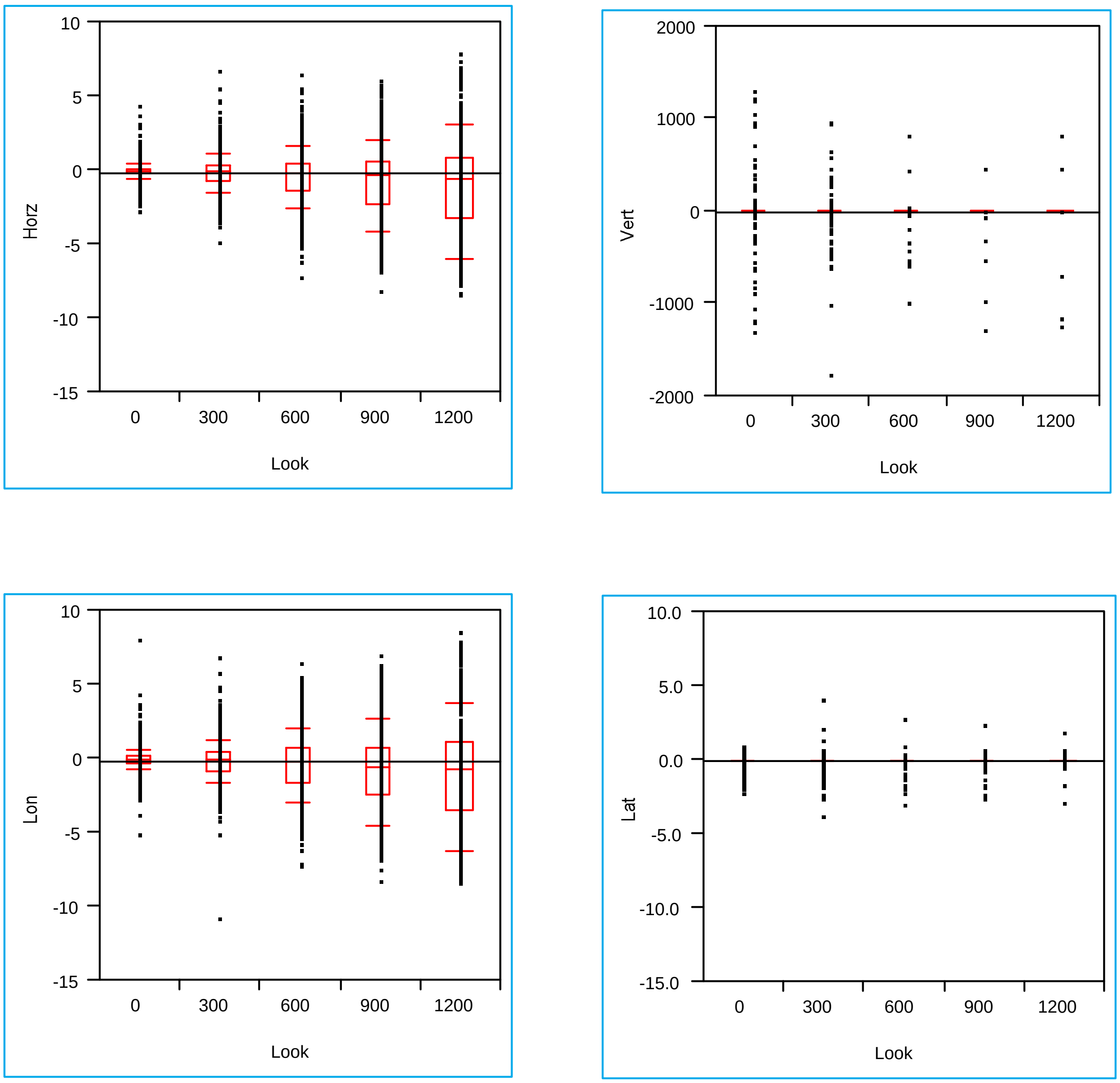


Figure A1- 1 Quantile Plots of Error Measurement for Run 000-100 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

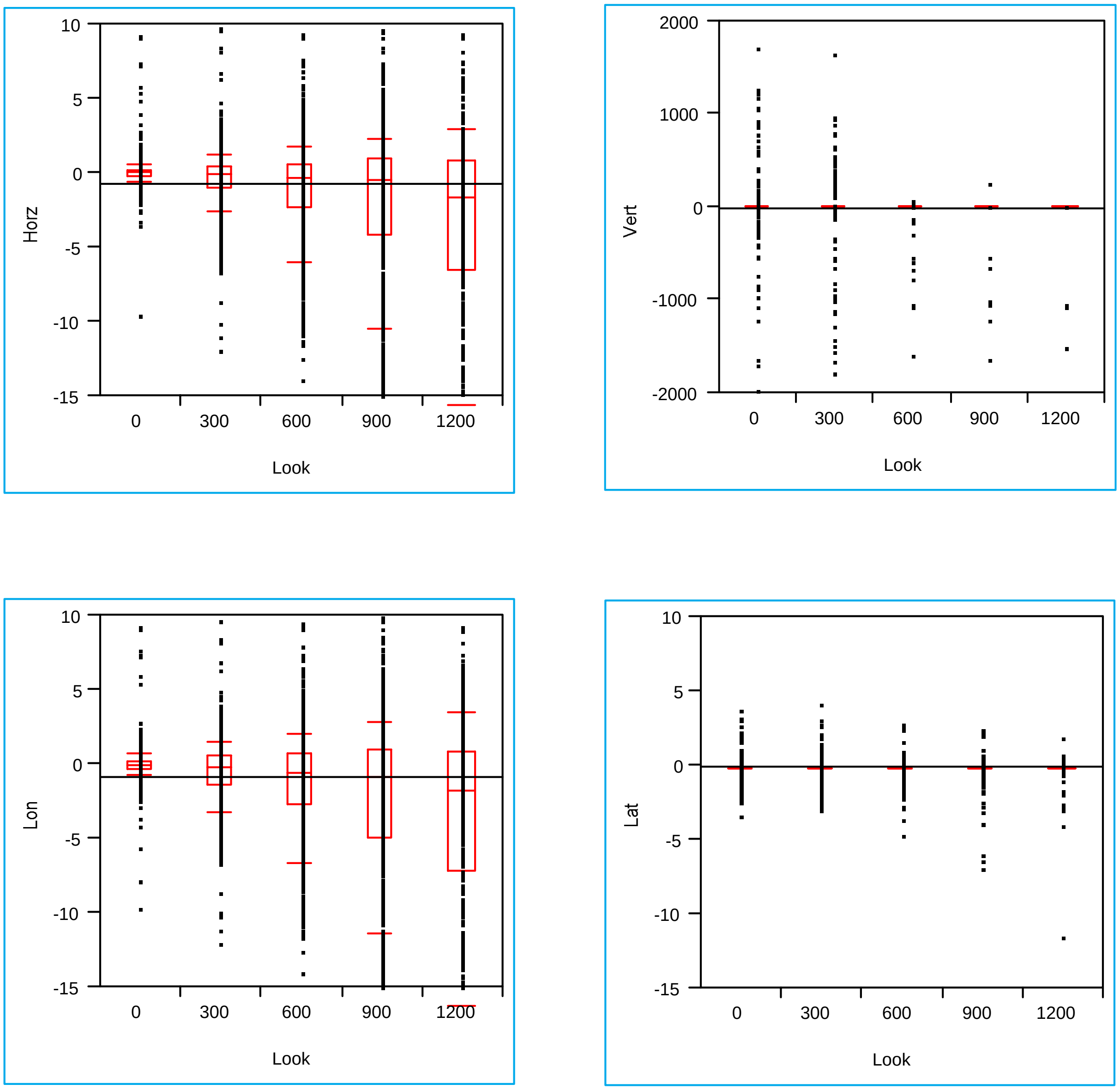


Figure A1- 2 Quantile Plots of Error Measurement for Run 000-200 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

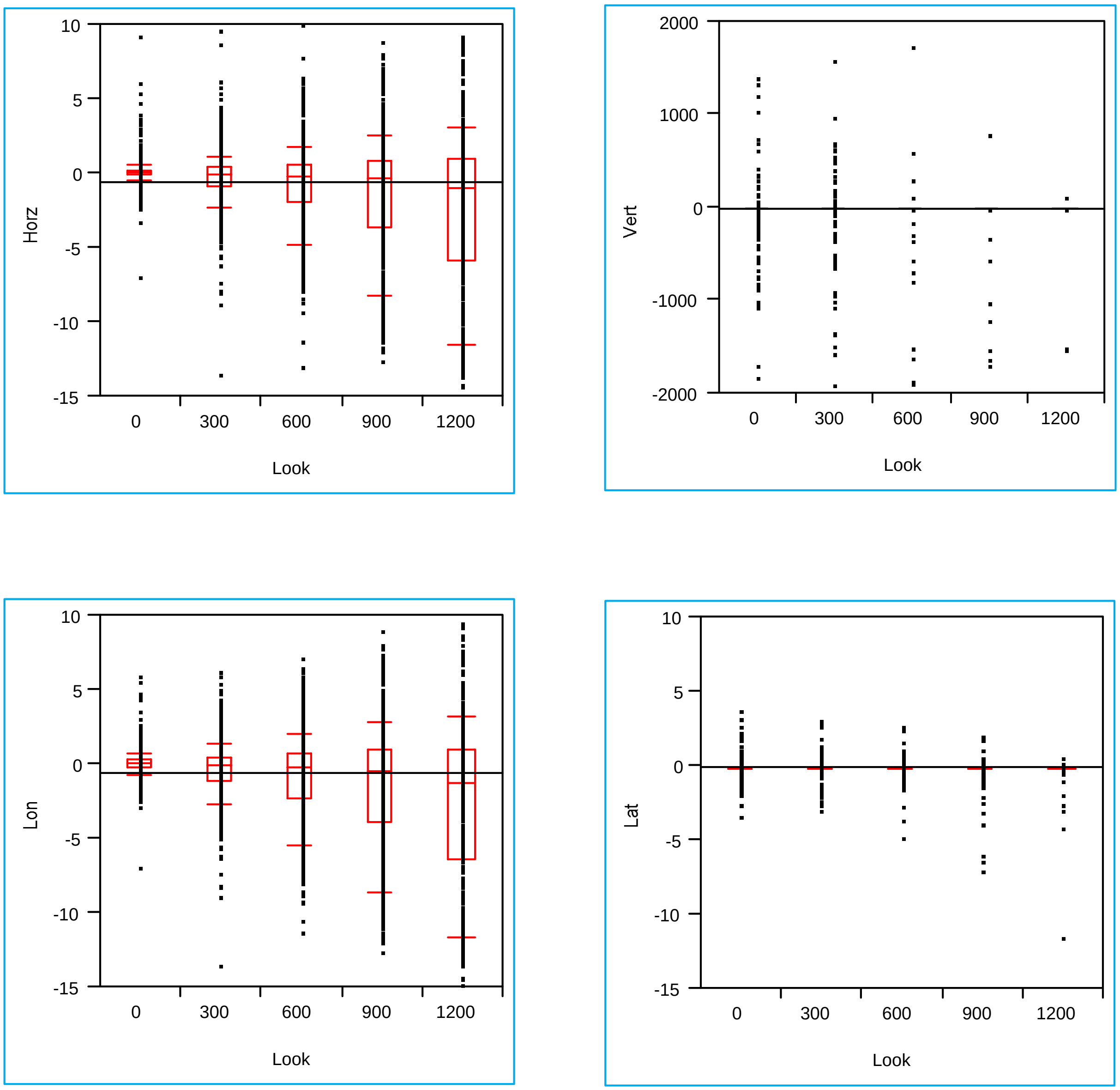


Figure A1- 3 Quantile Plots of Error Measurement for Run 100-200 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

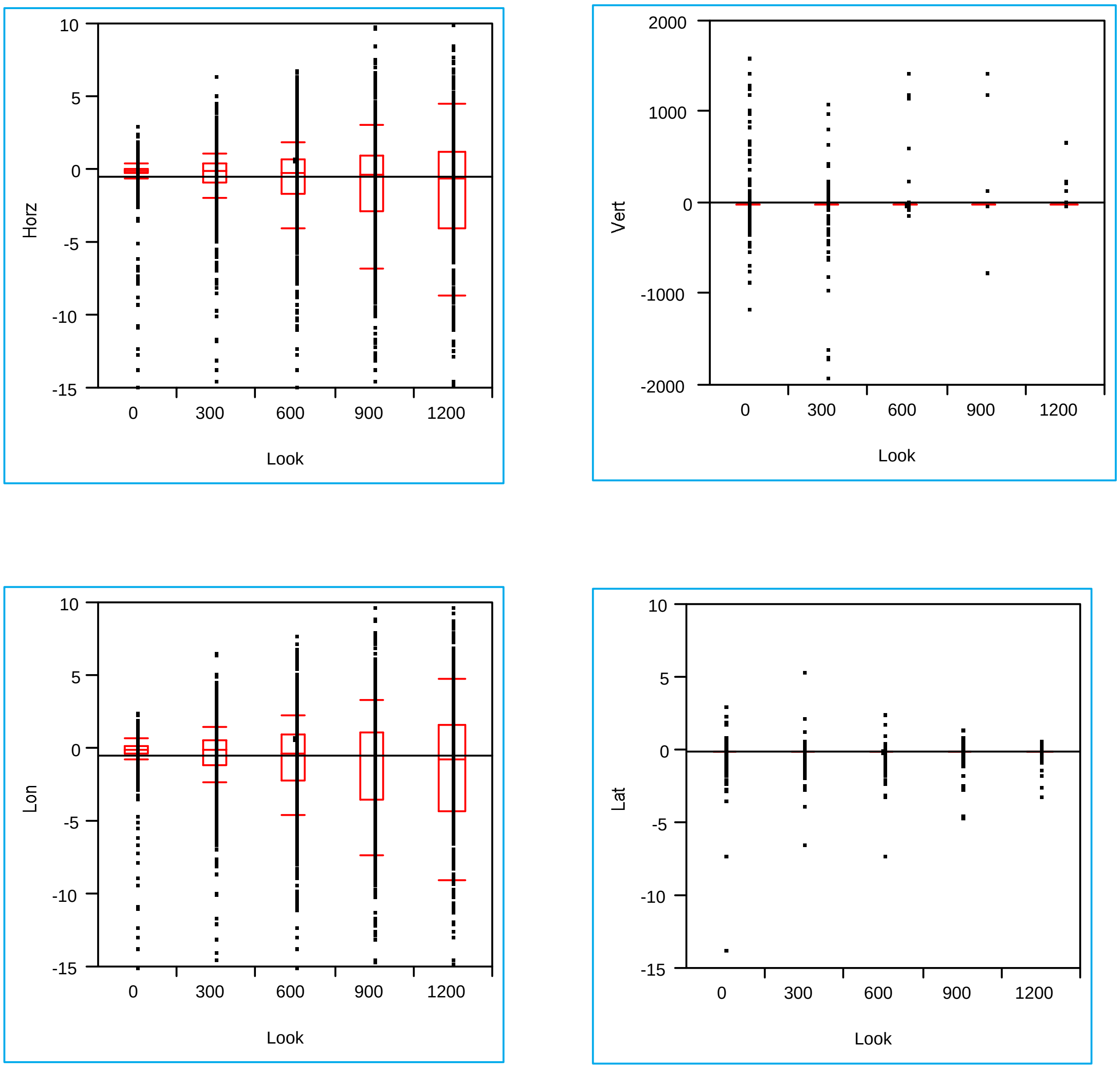


Figure A1- 4 Quantile Plots of Error Measurement for Run 000-010 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

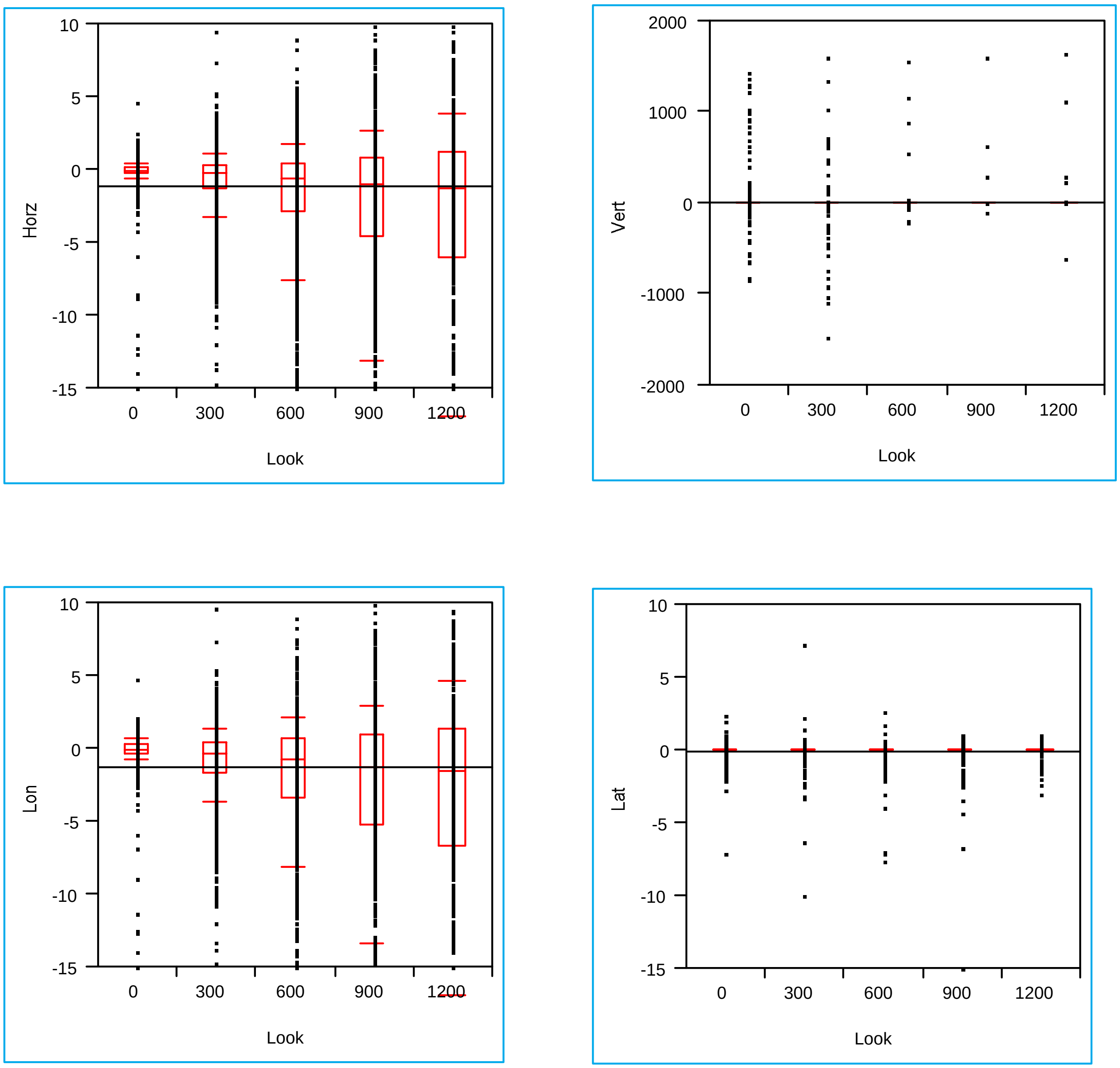


Figure A1- 5 Quantile Plots of Error Measurement for Run 000-020 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

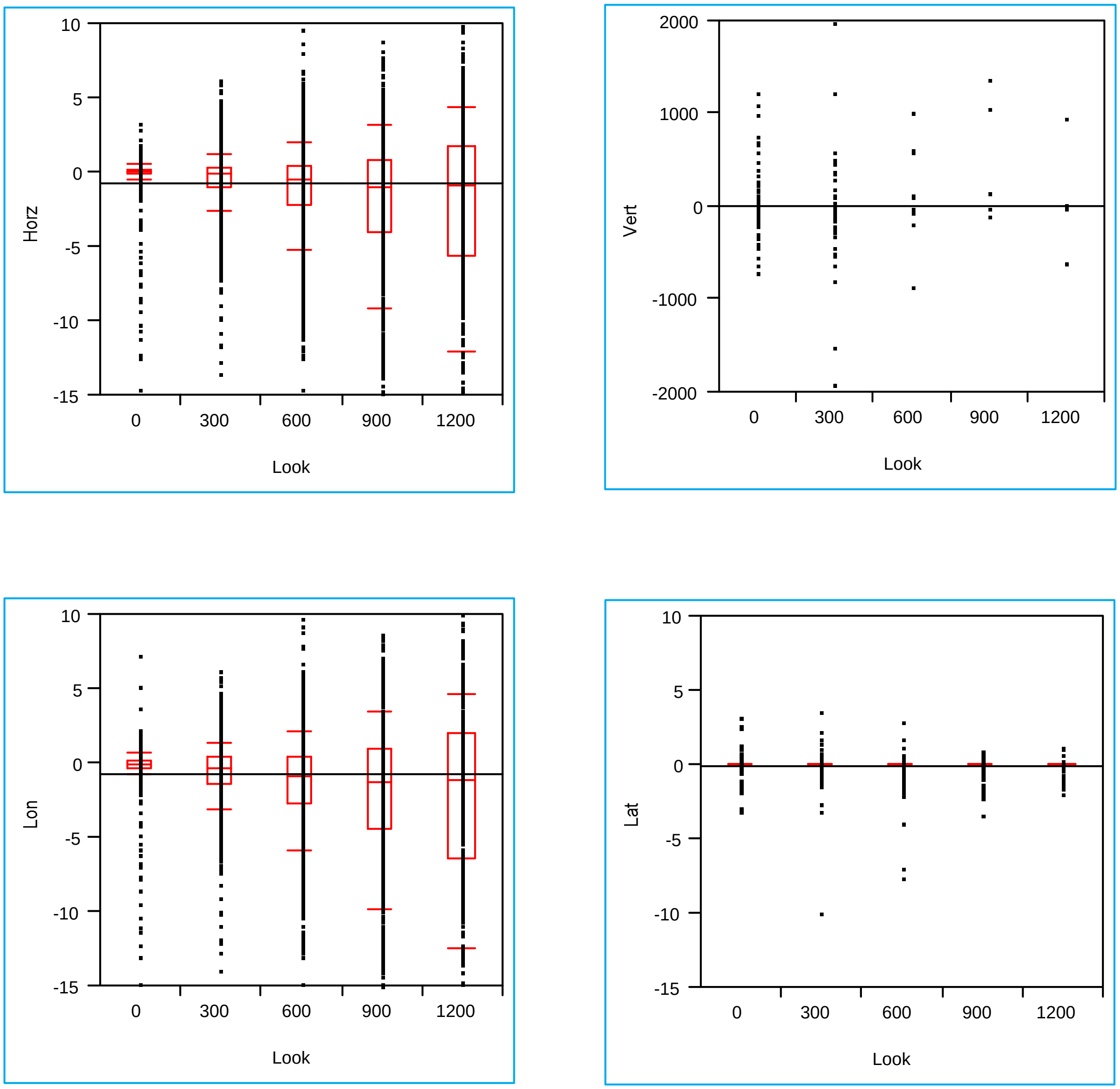


Figure A1- 6 Quantile Plots of Error Measurement for Run 010-020 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

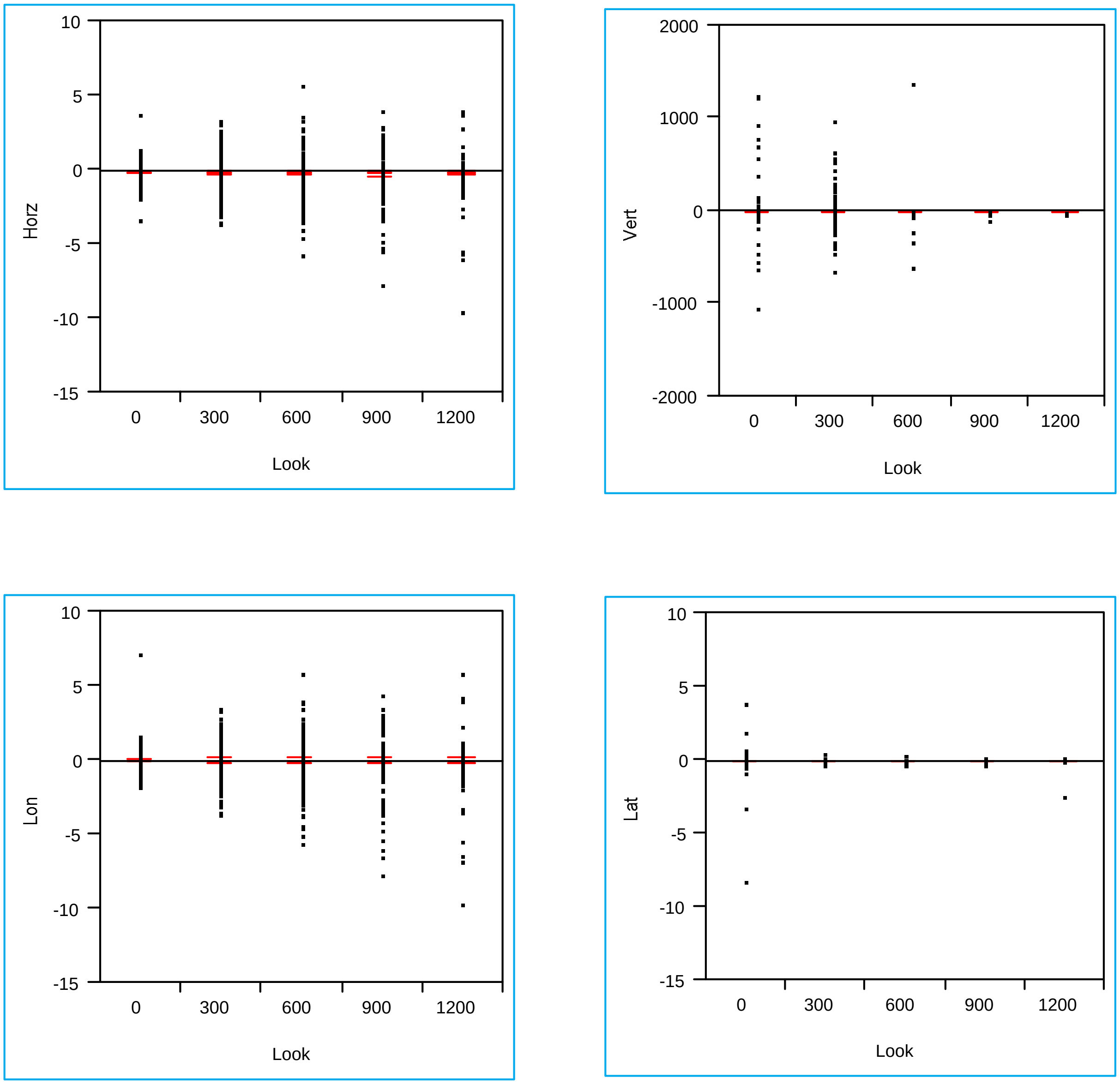


Figure A1- 7 Quantile Plots of Trajectory Error Measurement for Run 000-001 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

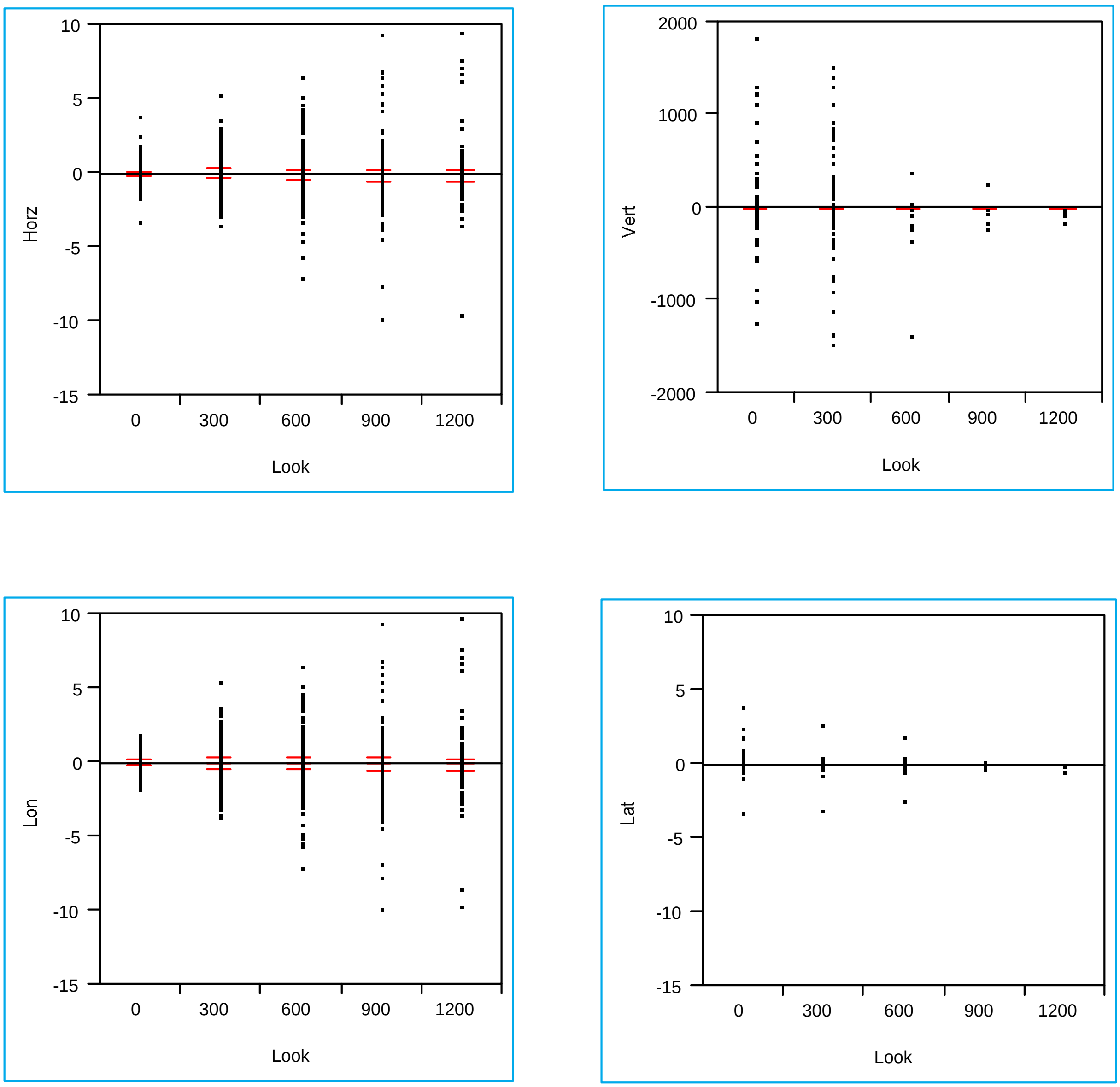


Figure A1- 8 Quantile Plots of Trajectory Error Measurement for Run 000-002 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

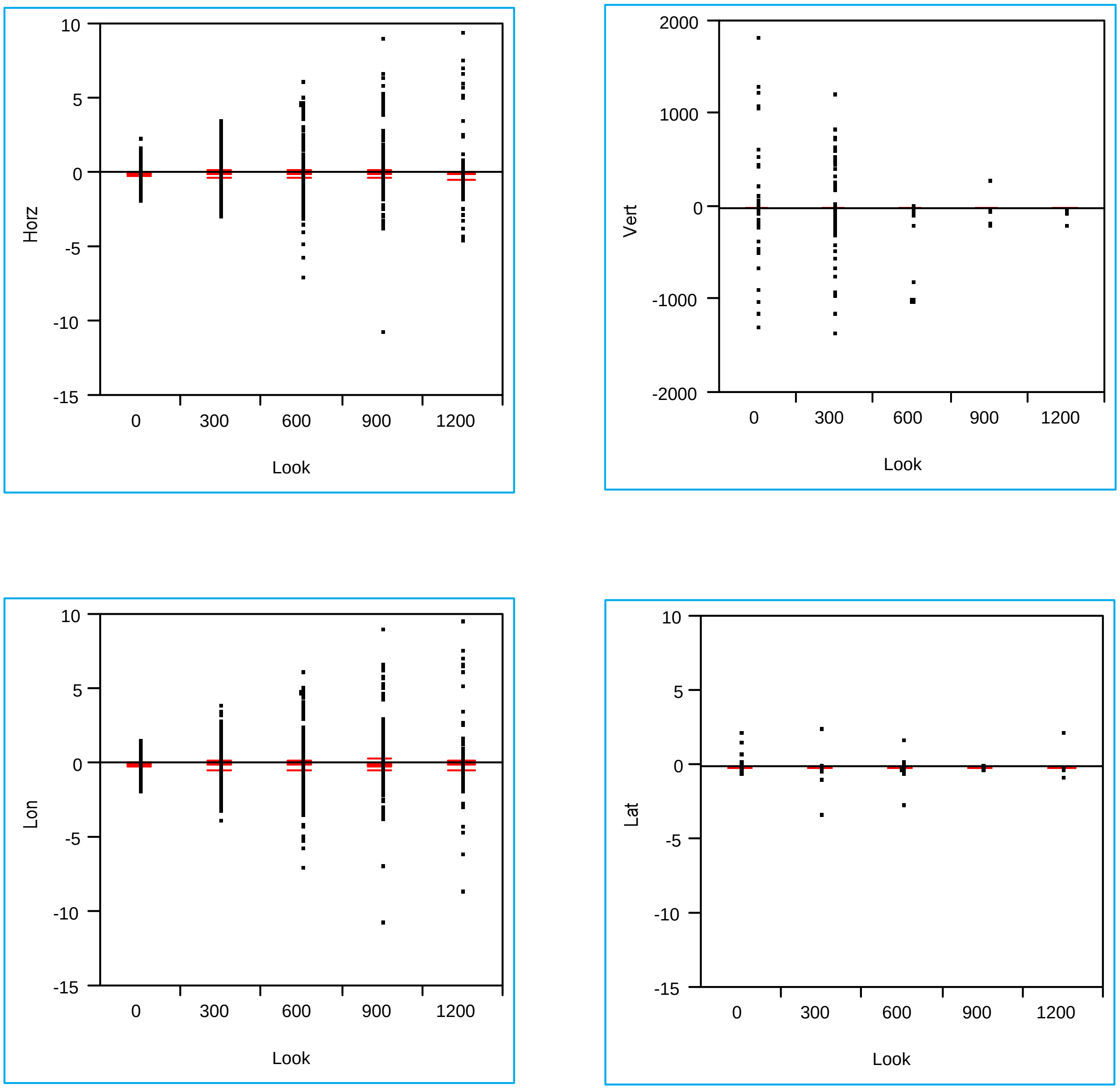


Figure A1- 9 Quantile Plot of Error Measurement for Run 001-002 with Level Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

A.1.2 JMP Box Plots for In-Transition Flight

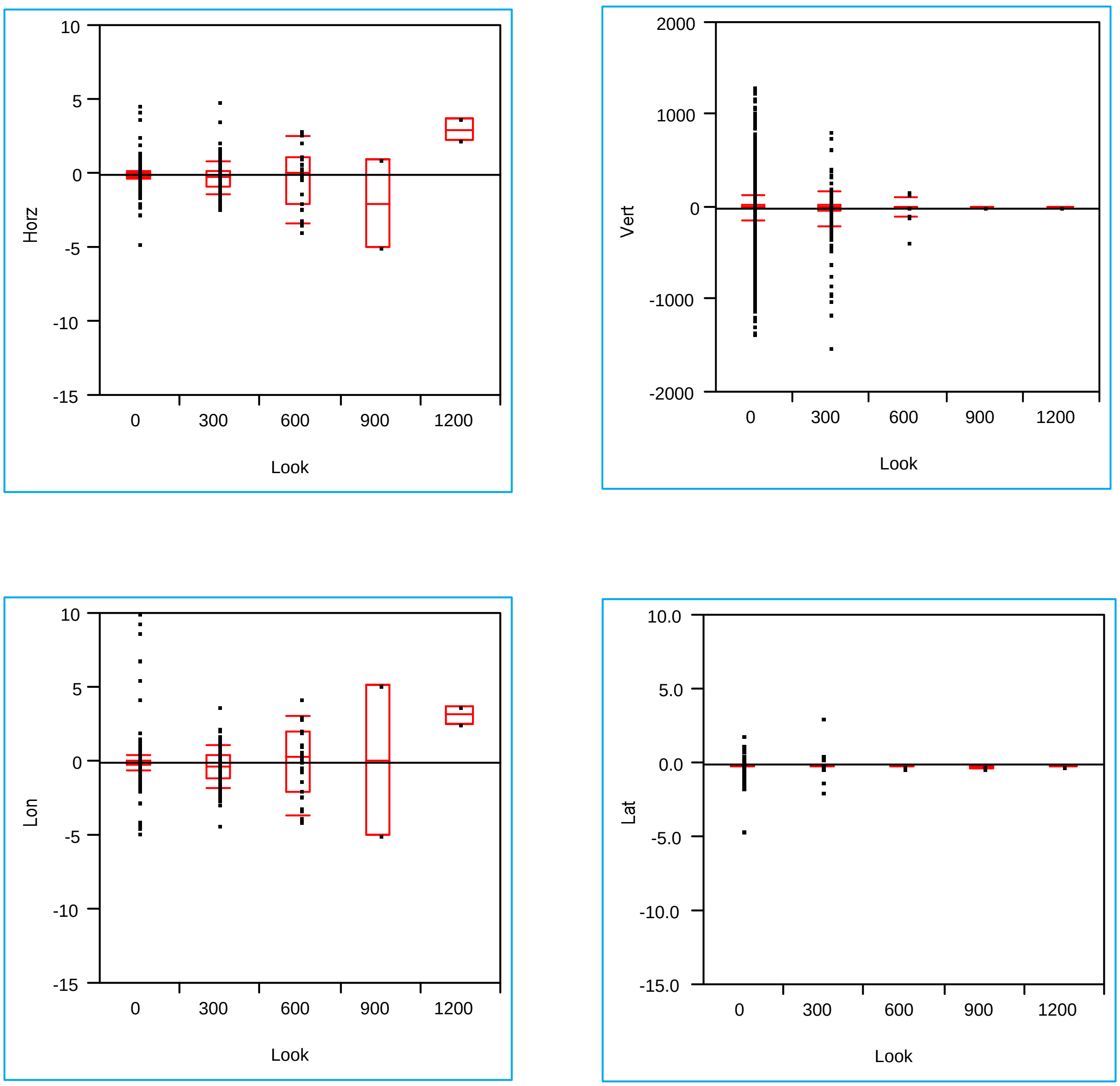


Figure A1- 10 Quantile Plots of Error Measurement for Run 000-100 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

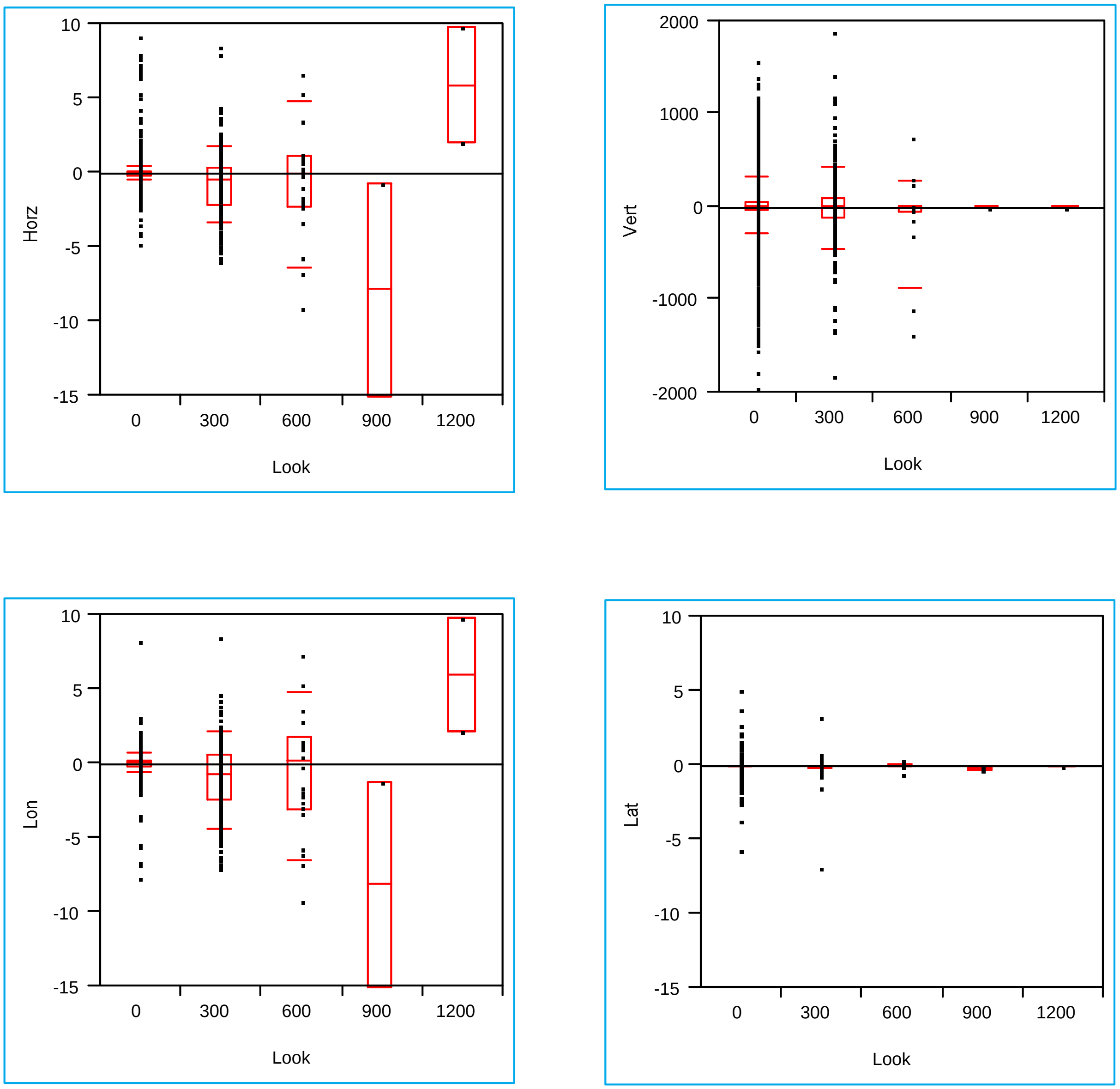


Figure A1- 11 Quantile Plots of Error Measurement for Run 000-200 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

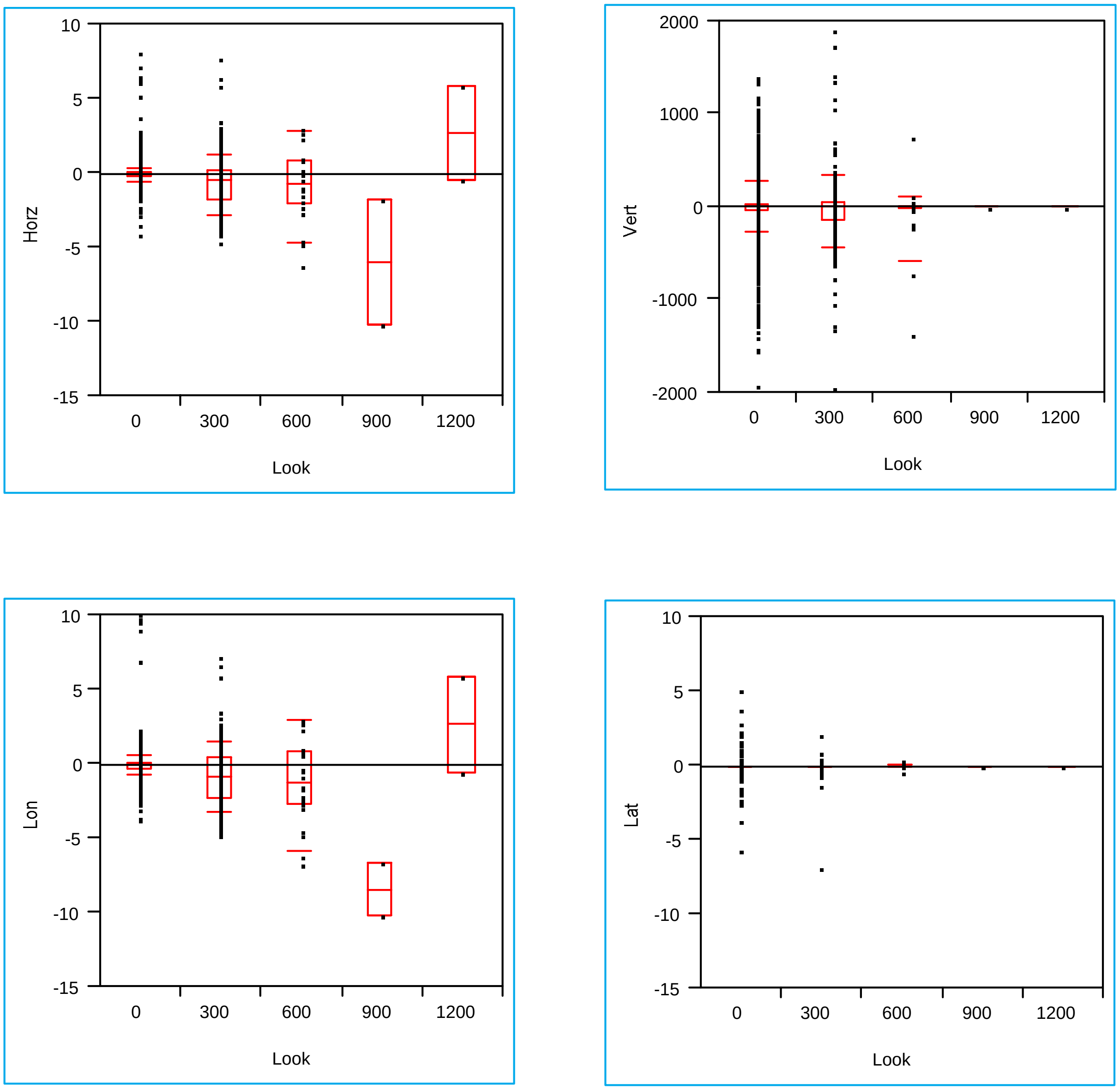


Figure A1- 12 Quantile Plots of Error Measurement for Run 100-200 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

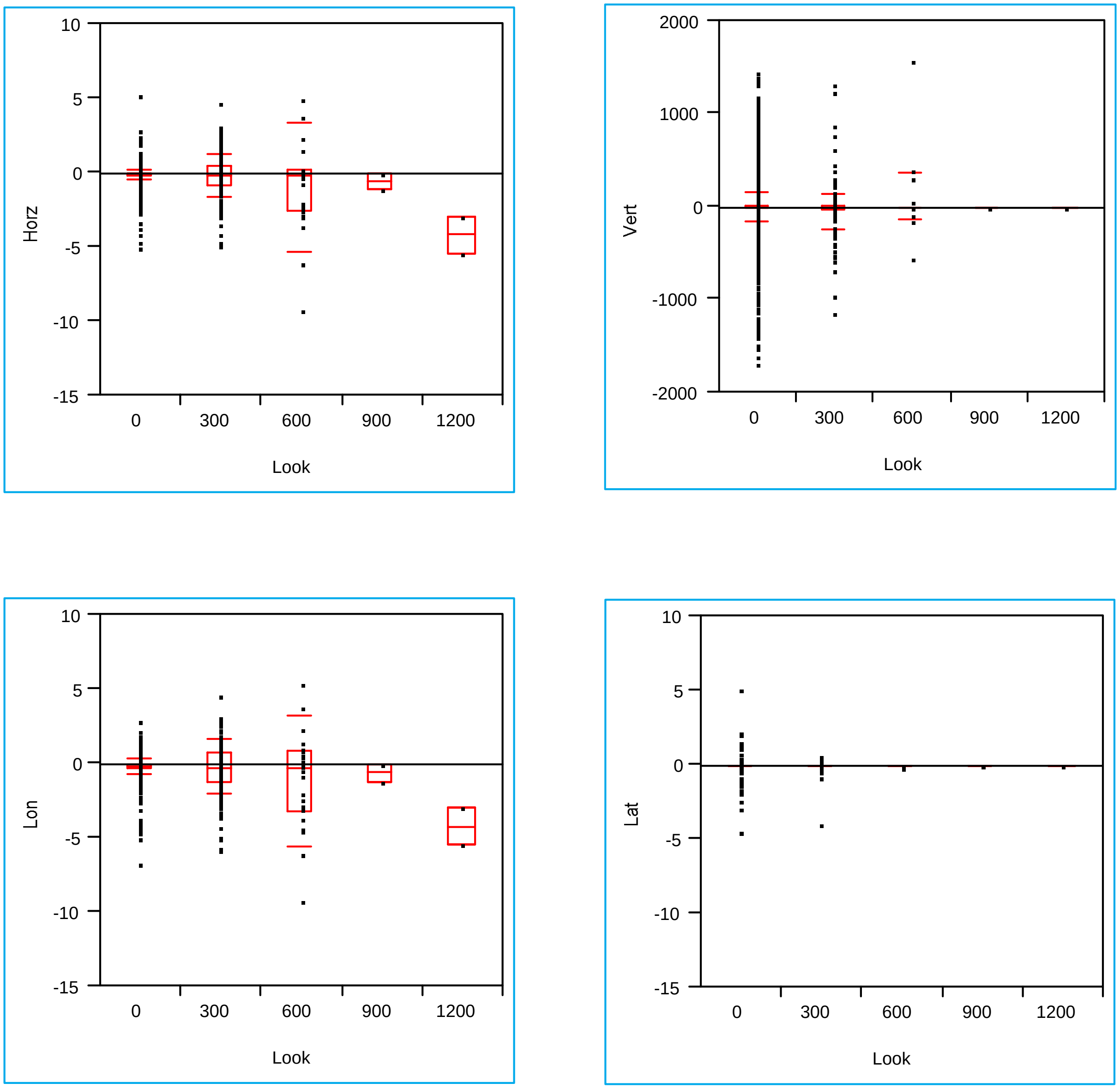


Figure A1- 13 Quantile Plots of Error Measurement for Run 000-010 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

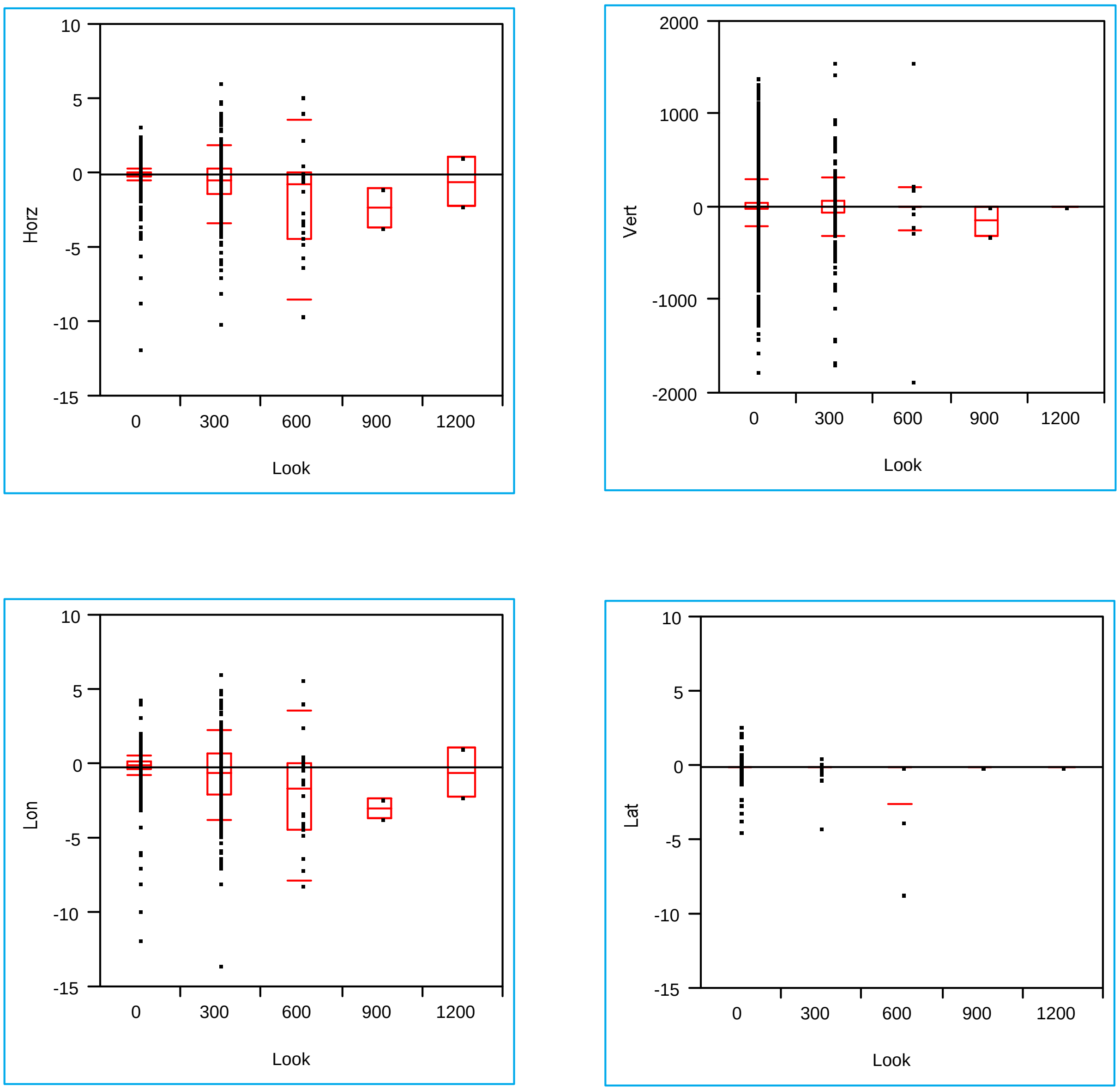


Figure A1- 14 Quantile Plots of Trajectory Error Measurement for Run 000-020 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

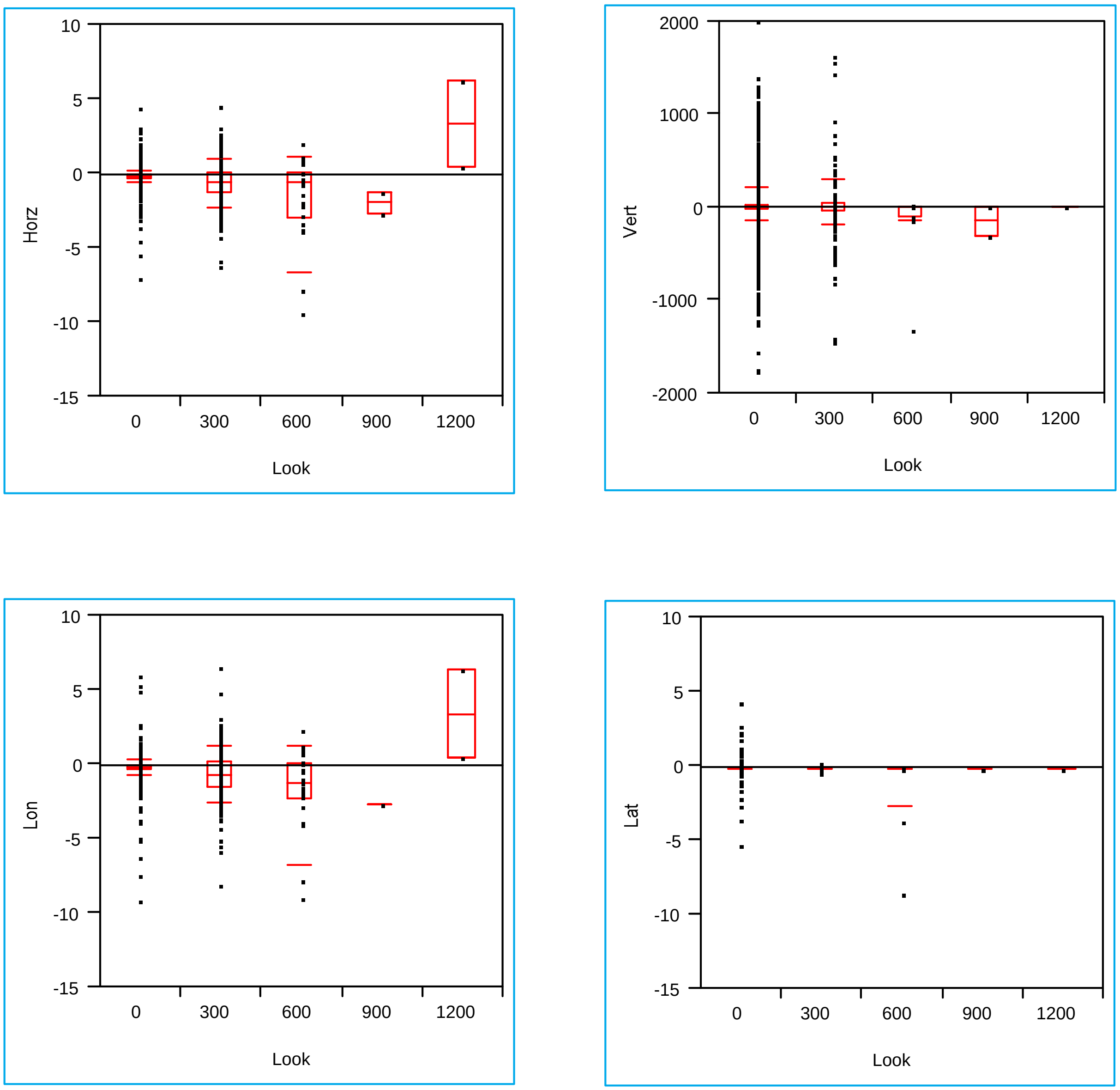
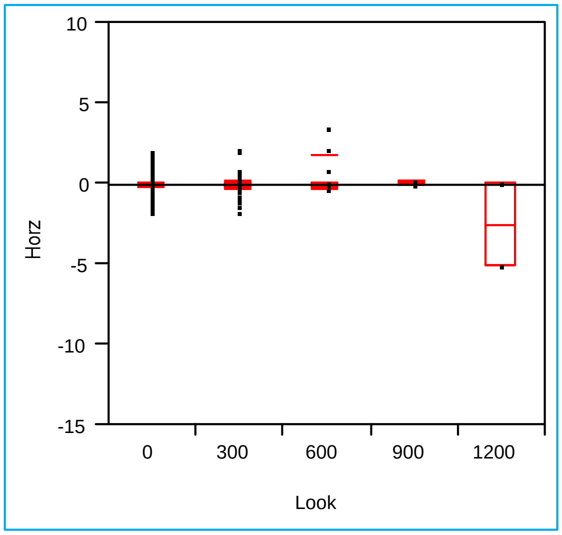
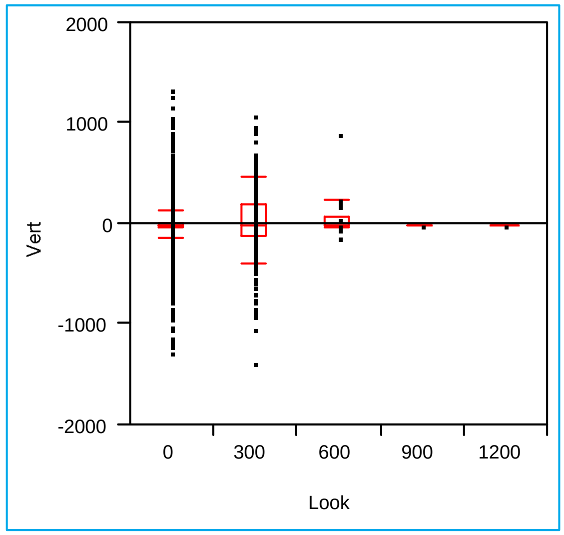
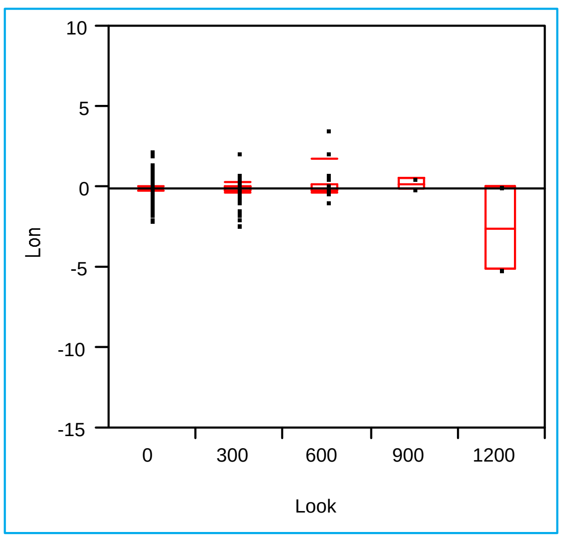


Figure A1- 15 Quantile Plots of Trajectory Error Measurement for Run 010-020 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)







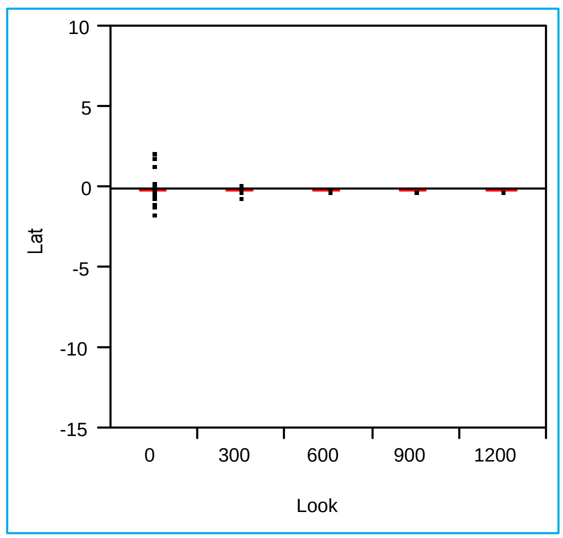


Figure A1- 16 Quantile Plots of Error Measurement for Run 000-001 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

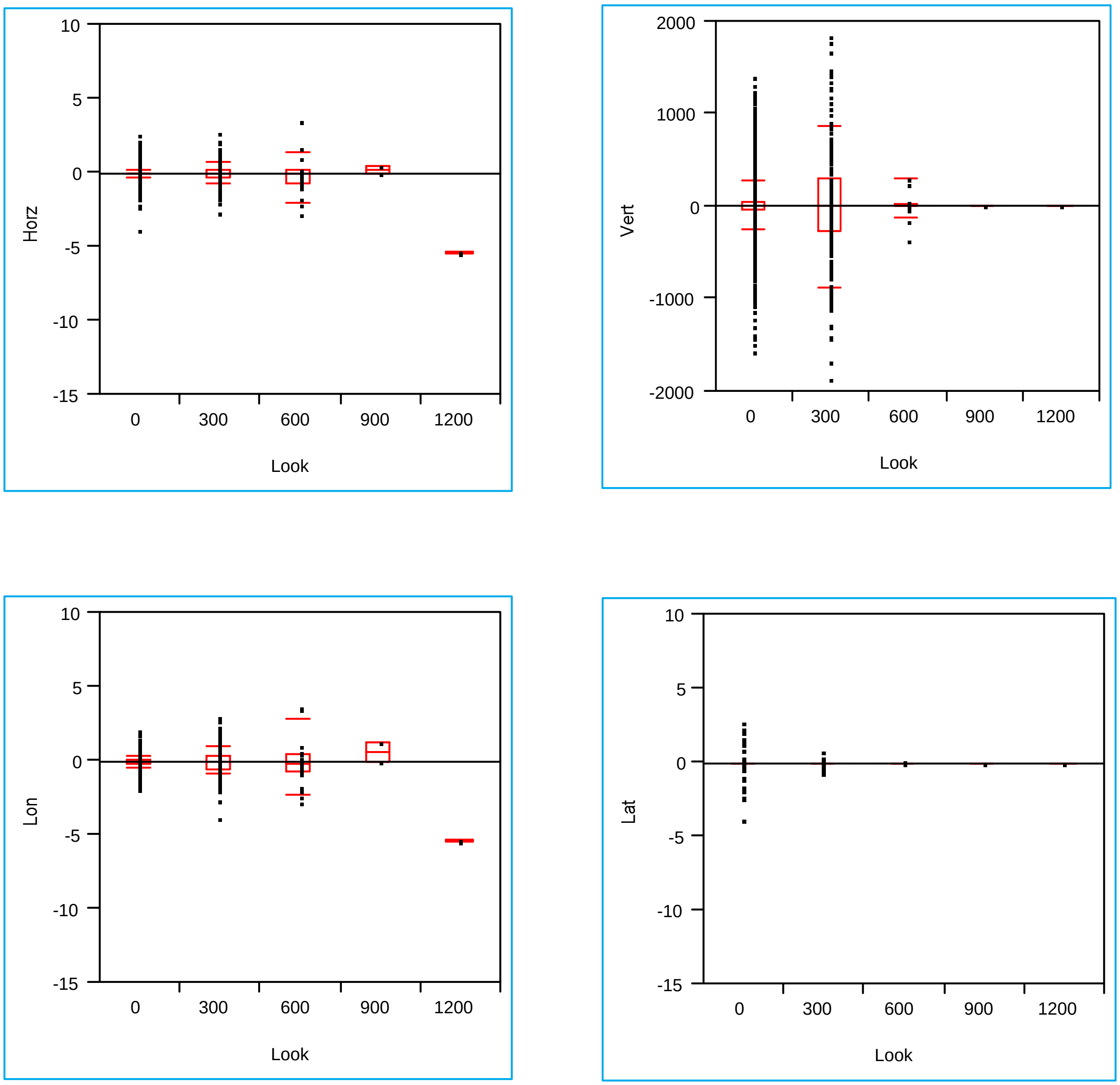


Figure A1- 17 Quantile Plots of Error Measurement for Run 000-002 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

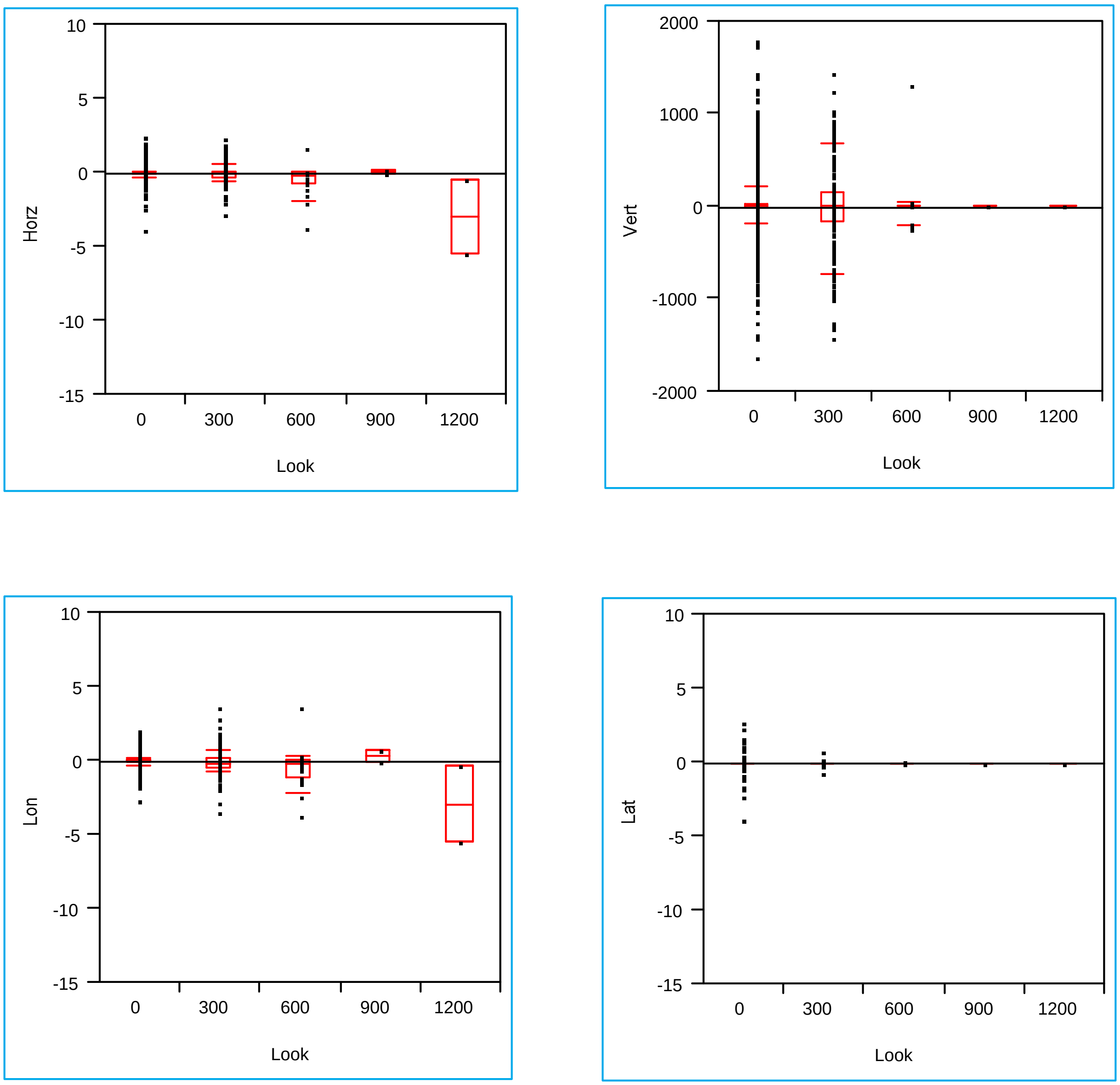


Figure A1- 18 Quantile Plots of Error Measurement for Run 001-002 with In-Transition Flight by Look Ahead Time (Horizontal, Vertical, Longitudinal, Lateral Error)

B. Appendix B - Median Plots

B.1.1 Plots for Level Phase of flight

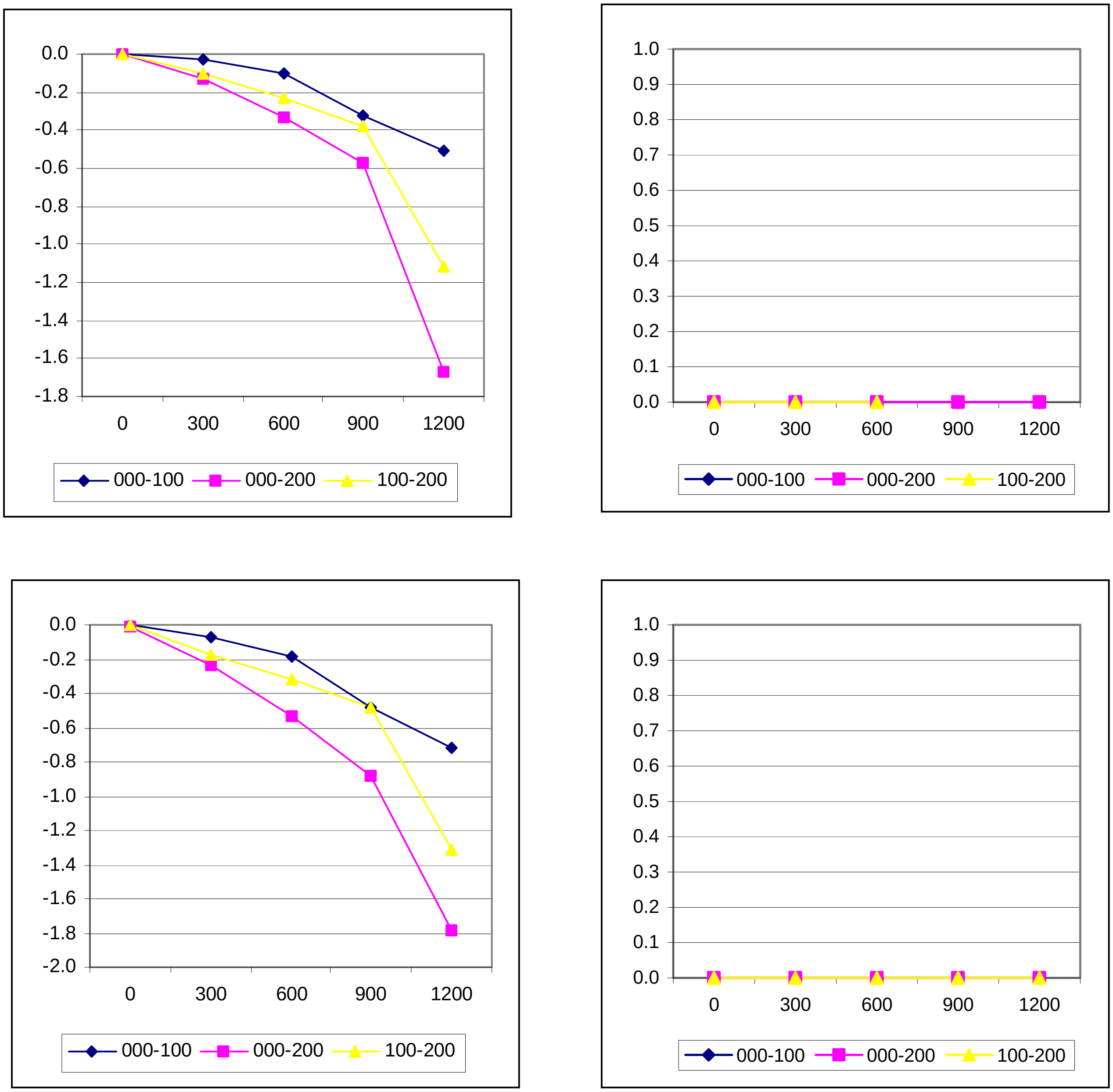


Figure B1- 1 Median Plots for Wind Magnitude (Horizontal, Vertical, Longitudinal, Lateral Error)

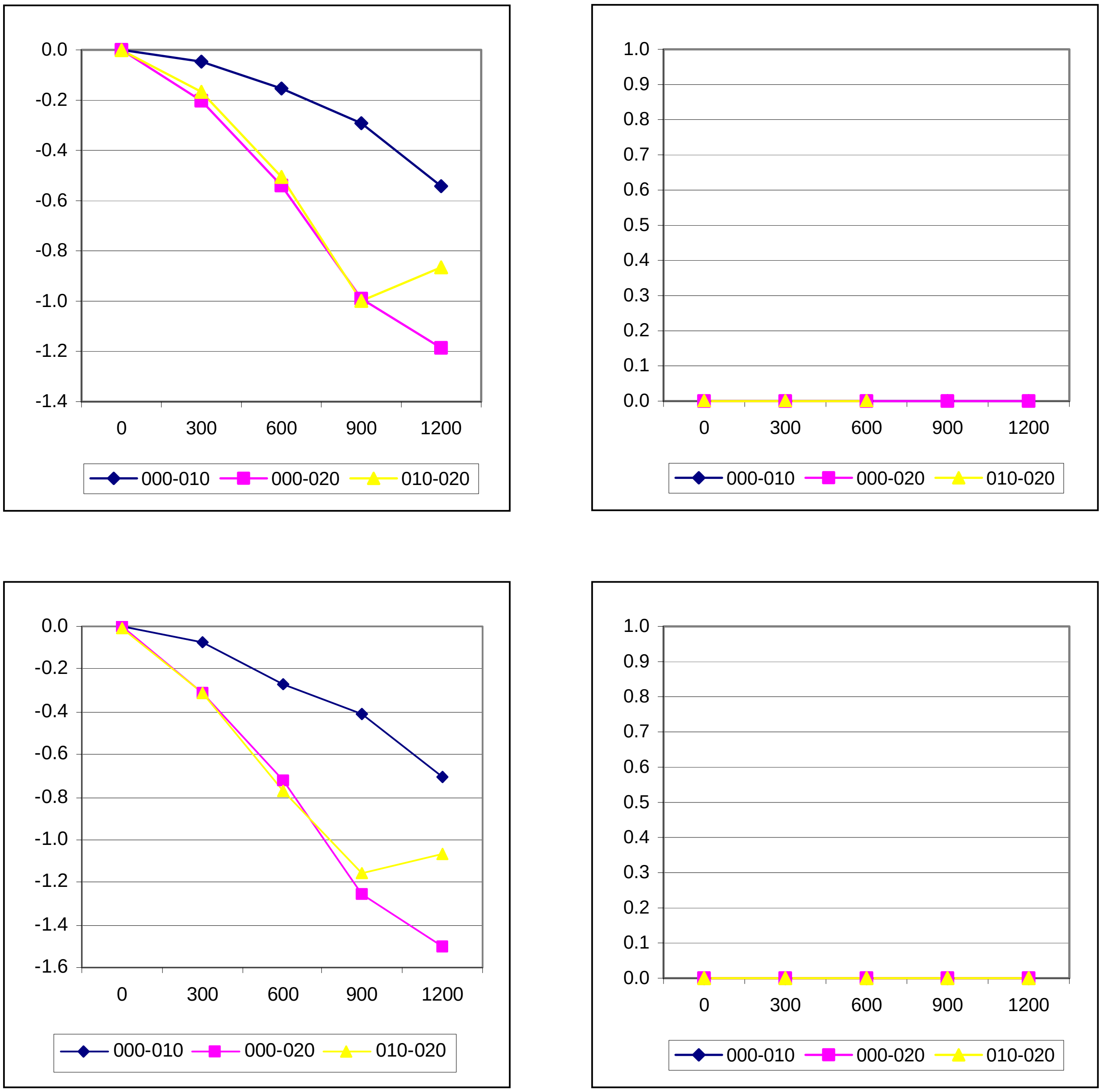


Figure B1- 2 Median Plots for Wind Direction (Horizontal, Vertical, Longitudinal, Lateral Error)

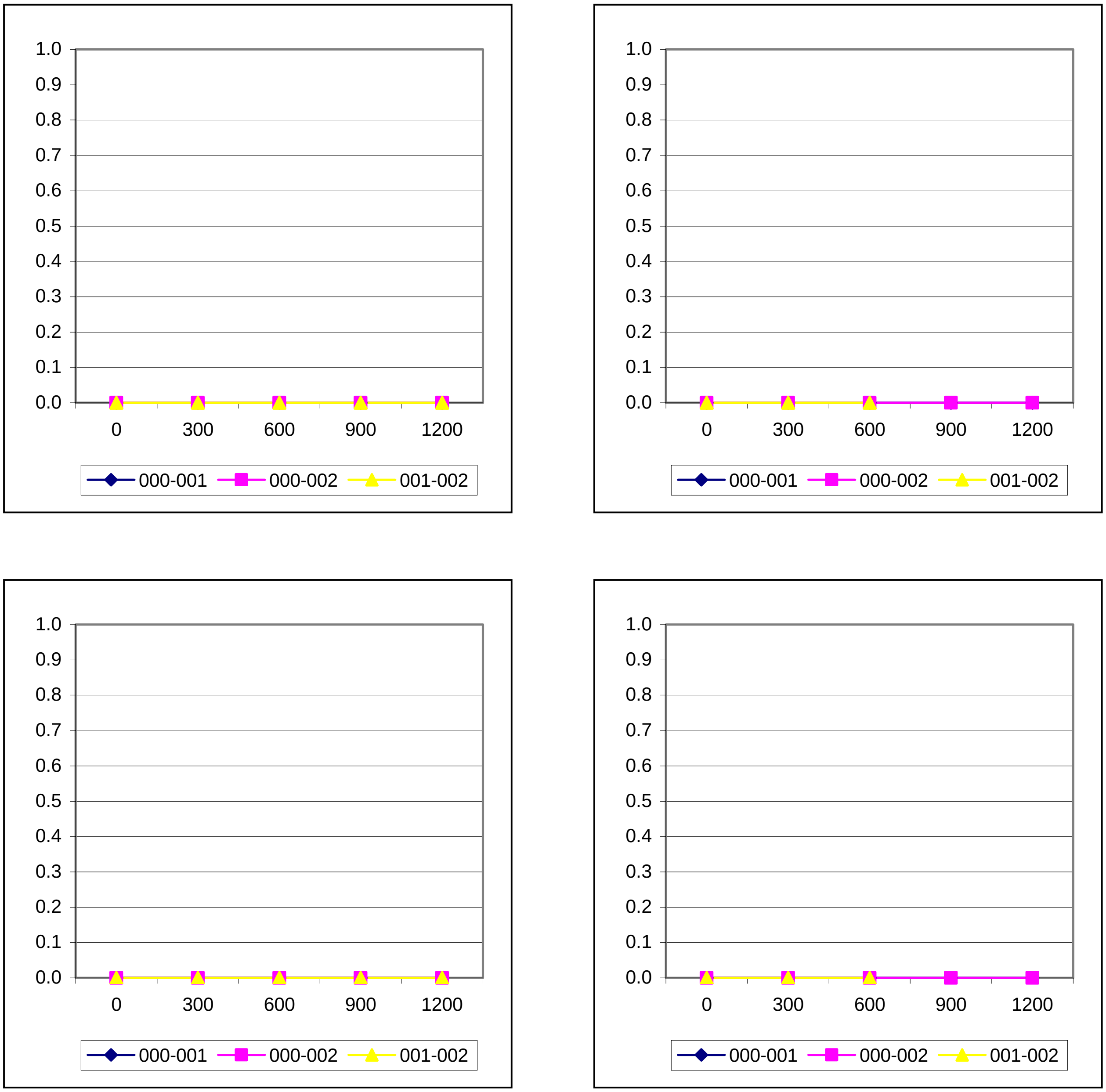


Figure B1- 3 Median Plots for Air Temperature (Horizontal, Vertical, Longitudinal, Lateral Error)

B.1.2 Median Plots for In-Transition Phase of Flight

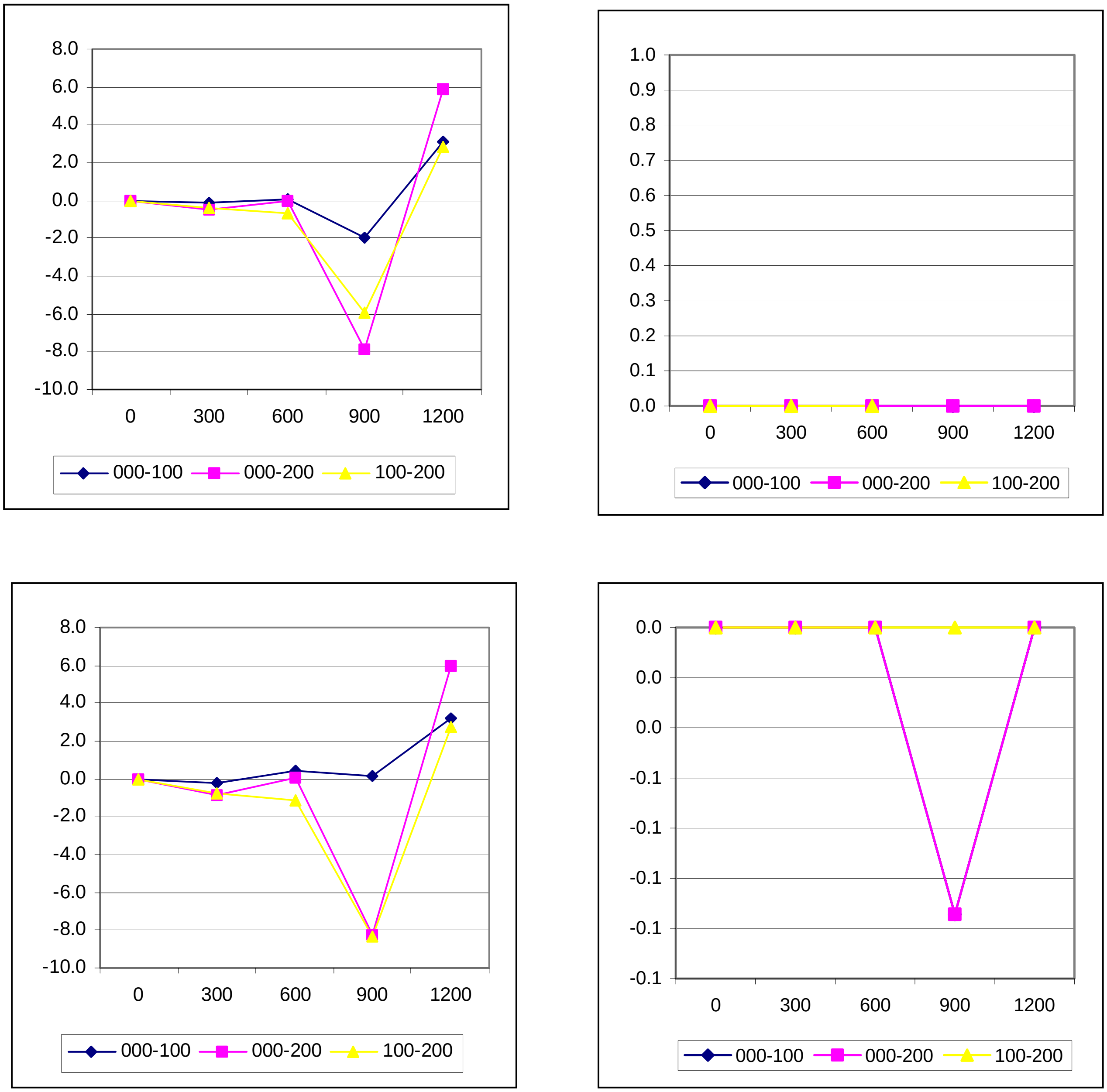


Figure B1- 4 Median Plots for Wind Magnitude (Horizontal, Vertical, Longitudinal, Lateral Error)

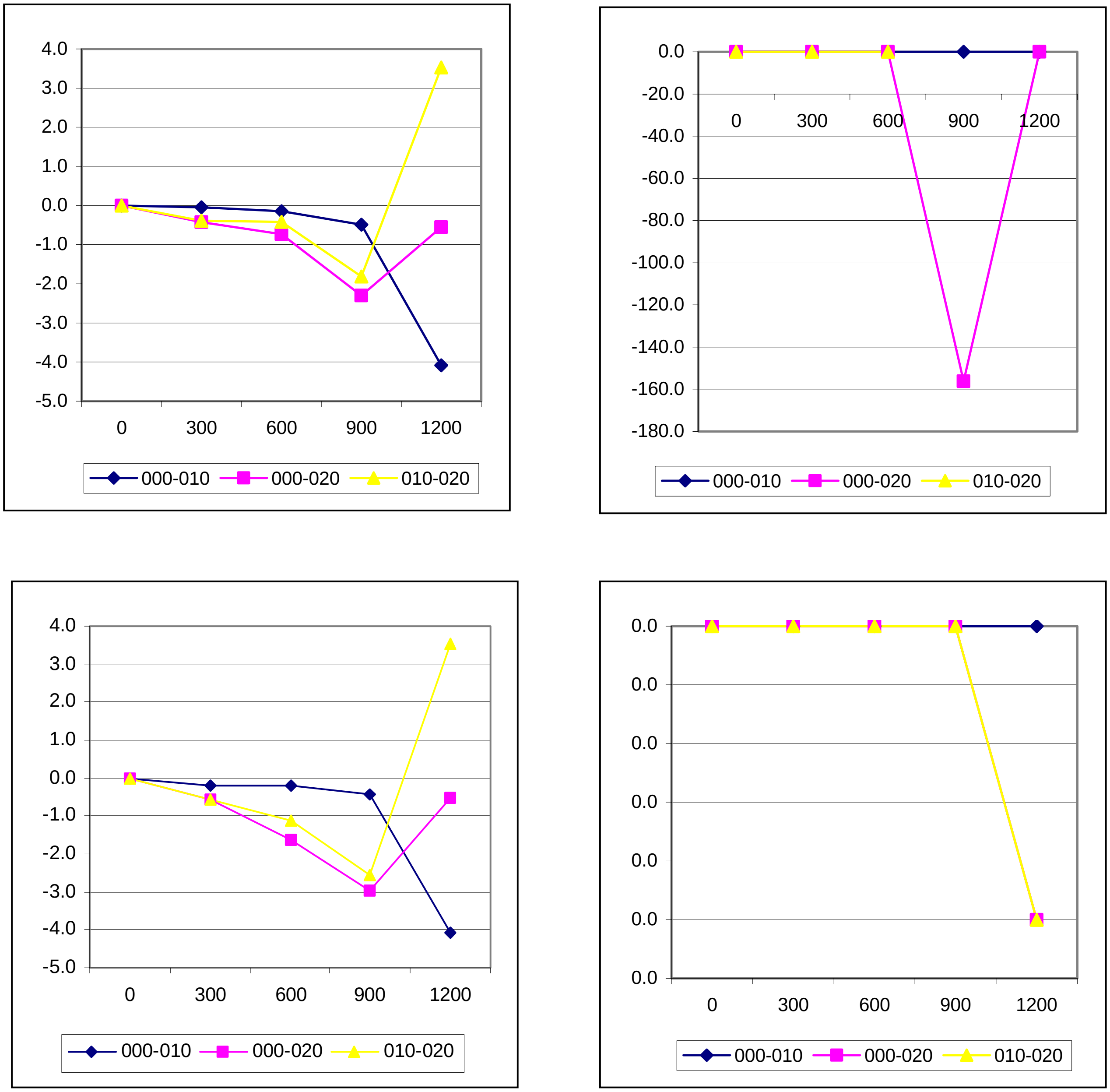


Figure B1- 5 Median Plots for Wind Direction (Horizontal, Vertical, Longitudinal, Lateral Errors)

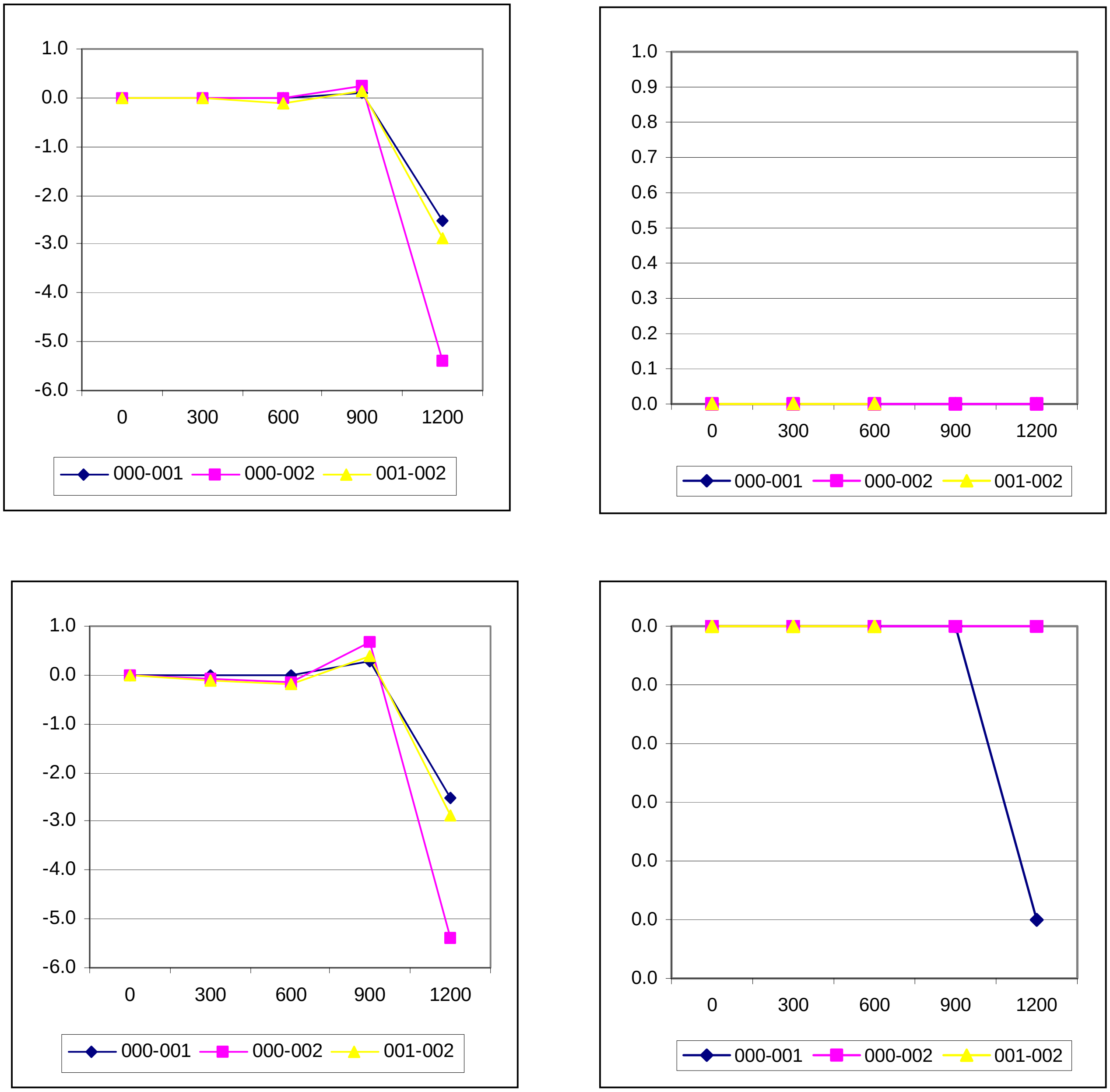


Figure B1- 6 Median Plots for Air Temperature (Horizontal, Vertical, Longitudinal, Lateral Error)

C. Appendix C - Preliminary Wind Data Analysis

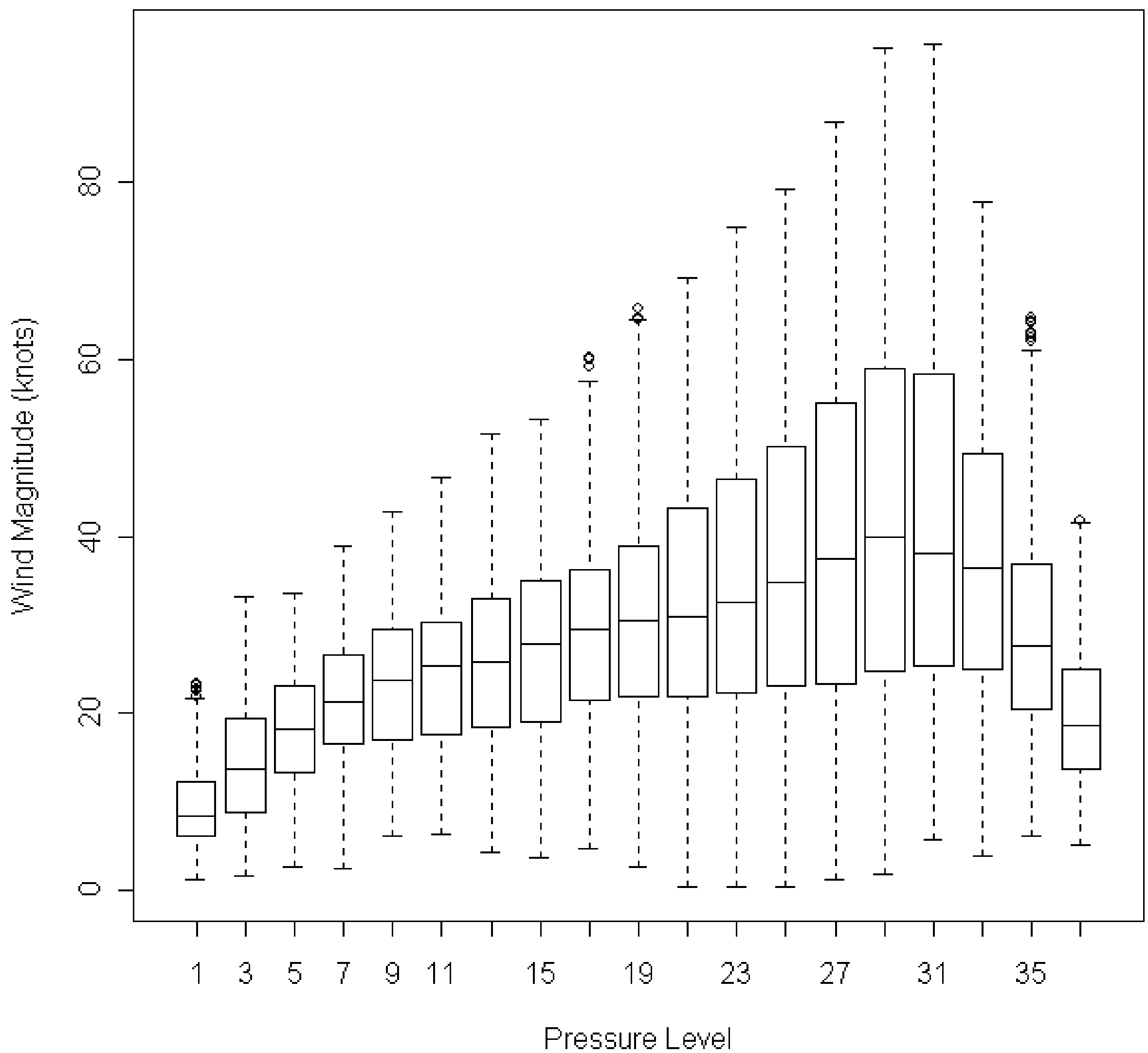


Figure C1- 1 Box Plot of Wind Magnitude Data by Pressure Level for ZID Center (05/26/1999)

[THIS PAGE IS INTENTIONALLY LEFT BLANK]

D. Appendix D – Detailed Flight Example

TBS